Uniting Big and Little Data to Understand Visitor Behavior

Jessica Roberts
University of Illinois at Chicago
jrober31@uic.edu

Amartya Banerjee
Northwestern University
amartya@u.northwestern.edu

Matthew Matcuk
Field Museum of Natural History
mmatcuk@fieldmuseum.org

Steven McGee
Northwestern University
s-mcgee@northwestern.edu

Michael Horn
Northwestern University
michael-horn@northwestern.edu

Introduction

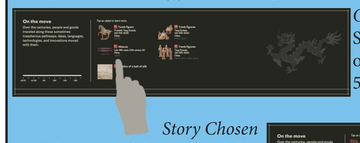
Traditional methods for understanding how visitors engage with exhibit elements are limited in their ability to find out how visitors are using digital interactives. Data logs from these interactives can provide aggregate information about overall usage, but in-person tracking is necessary to make sense of individual behaviors. In order to gain a full understanding of how interactives are impacting visitor experiences, both types of data must be analyzed and combined to mutually inform each other. Ongoing research combining Google Analytics data from touchscreen interactives, in-person timing and tracking observations, and tracking data from a custom-built mobile app that traces individual visitor paths through an exhibition is working to advance our understanding of how Big Data and little data can be used to better understand the visitor experience.

Digital rails throughout the China Hall provide interpretive information for all 350 objects displayed in the exhibition. Visitors can select from up to three stories about each object or group of objects. Some stories include supplemental digital materials consisting of over 1,100 zoomable images, slideshows, videos, custom maps, and animations, including 53 360-degree object viewers.

Two Data Types

Google Analytics

Data from all digital rails were pulled for the entire 3-week period of the timing and tracking study, totaling over 200,000 unique events logged. Examples:



Object Chosen
Select one of the 6-8 available objects or group of objects.
55,433 events

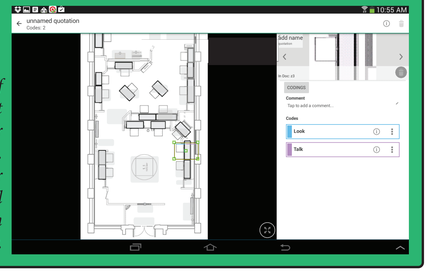
Story Chosen
Select one of the 2-3 available stories for an object or group.
83,247 events

Slide Reached
Hit one of the slides in a slideshow.
32,657 events

Events were grouped by rail, with an average of 4,653 events per rail and a standard deviation of 1,768.

In-Person Timing and Tracking

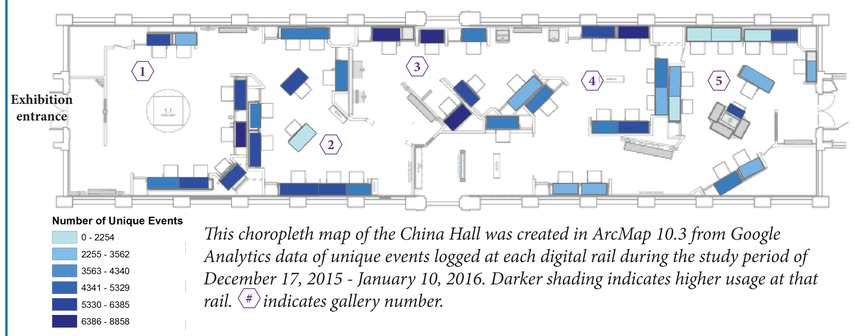
An uncued timing and tracking study was conducted over 3 weeks in December 2015-January 2016. When a group of visitors entered the main door of the China Hall, a researcher used a tablet application to record behaviors such as looking at an object case or rail, touching a rail or element, speaking with a companion or a docent, or taking a photo. These annotations were coded with location IDs and exported for analysis in Excel and ArcMap.



Screen capture of ATLAS.ti tablet application used for data collection. Annotations to the floor plan of the China Hall were tagged for each targeted behavior.

Analyzing and Visualizing Across Datasets

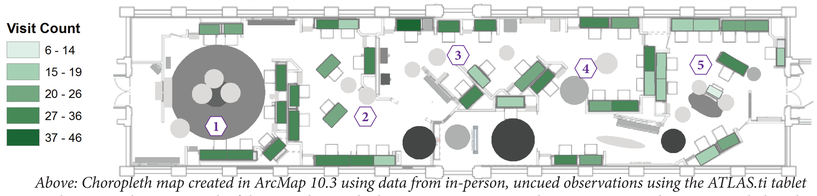
Google Analytics: Big Data to show large trends



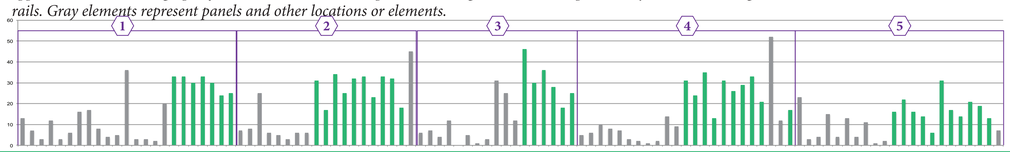
This choropleth map of the China Hall was created in ArcMap 10.3 from Google Analytics data of unique events logged at each digital rail during the study period of December 17, 2015 - January 10, 2016. Darker shading indicates higher usage at that rail. (#) indicates gallery number.

Log files can give insight to popularity of objects but can't answer other important questions about visitor behaviors, such as when they engage with non-interactive objects or read information on a screen without selecting it (i.e. when the screen has not timed out after a previous user's interaction).

Visitor Tracking: Complementing and refining understandings

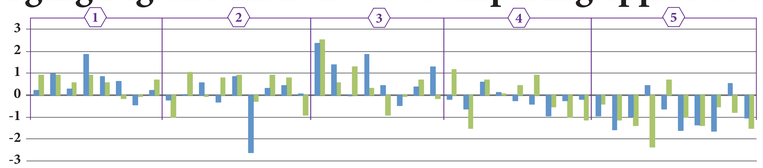


Above: Choropleth map created in ArcMap 10.3 using data from in-person, uncued observations using the ATLAS.ti tablet application. Below: Bar graph of the same data. In both representations, green elements represent object cases with digital rails. Gray elements represent panels and other locations or elements.



Data from in-person tracking allows us to see what other exhibit elements are holding visitors' attention, such as text panels and other visual displays but can't answer questions about individual touchscreen interactions.

Merging Big and Little Data: Comparing apples to apples



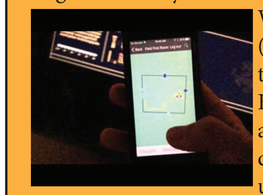
Standard scores were computed for the number of visits to each object case with a digital rail in the live tracking data (green) and number of events per rail logged by Google Analytics (blue). The y-axis represents standard deviations above and below the mean. Note the outlier in gallery 2: the rail nearly three standard deviations below the mean was broken for much of the study period and therefore had no log data from that time. Live tracking observed some visitors stopping by the case to view the object.

Comparing tracking and log data lets us better understand visitor behavior, particularly when there is a mismatch: does a particular case draw people to the object but not to the rail content? Is a rail intriguing to those who find it but overlooked by most visitors?

New Tracking Methods

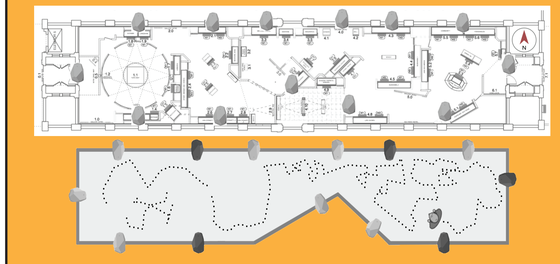
Bluetooth Beacons

Combining individual tracking data with log files can give valuable information on how visitors are using digital interactives over the course of their visit, but manually tracking visitors and synchronizing log files with tracking data is time-intensive and therefore cost-prohibitive for a large scale study.



We have installed Apple's iBeacon (Bluetooth low energy transceiver) technology throughout the China Hall to determine a visitor's approximate location. Proximity data from multiple beacons allows us to find an intersection (or common) area that corresponds to the indoor location of a visitor accurate within a 1-1.5 meter radius.

In ongoing visitor research, participants are asked to carry an iPod with a custom-built application enabling Bluetooth tracking. Paths recorded by this application are being compared to live timing and tracking by a researcher in order to assess the accuracy of the iBeacon technology.



Top: iBeacon placement within the China Hall.
Bottom: Sample trace of a visitor's path through the China Hall

Northwestern University

The Field Museum

We thank the National Science Foundation for their support of this project (IIS-1550051). Any opinions, findings, and recommendations are those of the authors and do not necessarily reflect the views of the NSF.