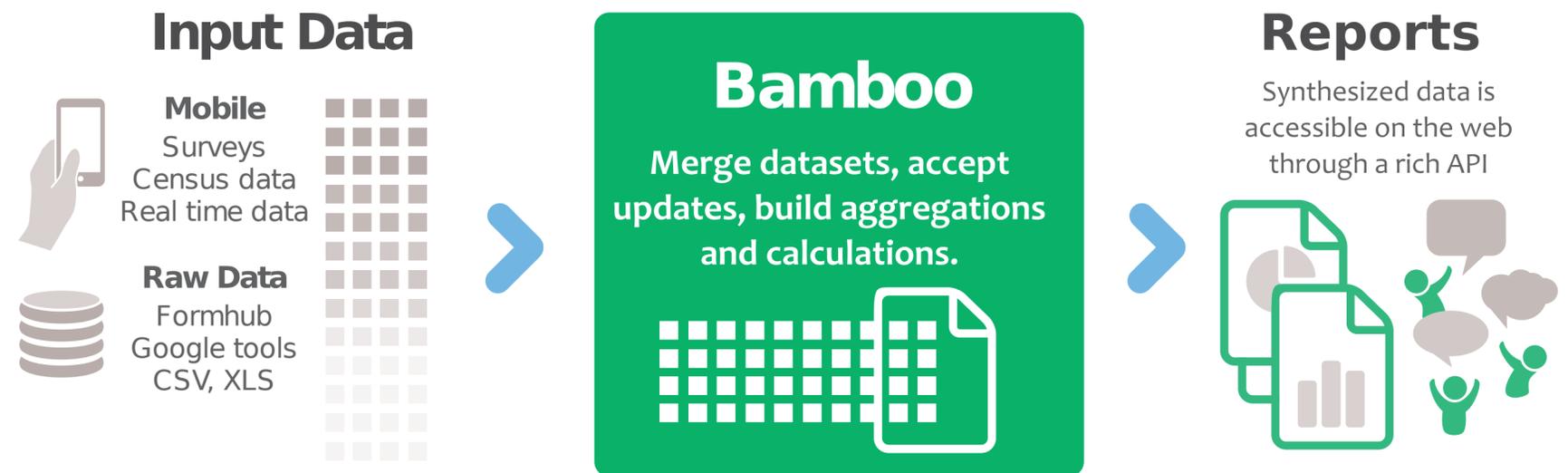# bamboo.io

Improved data collection and monitoring through dynamic data analysis

bamboo is a modular web service for real-time data analysis.

bamboo allows practitioners to analyze and aggregate datasets that update over time.

bamboo applies to facility inventories, summarization of patient data from community health workers, etc.

Input Data

Mobile
Surveys
Census data
Real time data

Raw Data
Formhub
Google tools
CSV, XLS

Bamboo

Merge datasets, accept updates, build aggregations and calculations.

Reports

Synthesized data is accessible on the web through a rich API

P. Lubell-Doughtie, P. Pokharel, M. Johnston, V. Modi

Implementation

- **Easy to use:** Users can upload CSV files and author formulas using a simple language.

- **Developer-friendly:** Updates are submitted to a JSON-based API which follows REST conventions; client libraries in python + javascript available.

- **High performance:** Uses pandas, a Python library for statistical data analysis; uses MongoDB for storage.

Monitoring water points with bamboo

Scenario: An NGO wants to monitor water points, by surveying them at a regular interval. At any point in time, staff want to know (1) when each water point was last surveyed and (2) whether it was functional.

1 **Connect data collection tool to bamboo,** e.g, with a single click on formhub.org.

2 **Author calculations** using bamboo's formula language.

3 **Retrieve and visualize aggregated data** on a web page, which will update as new data comes in.

bamboo vs. other tools

	Google Fusion Tables	R	**bamboo**
Connect to web-based data repositories	✓	[limited]	✓
Publish analyzed data to web	✓	–	✓
Pull in future data updates	✓	–	✓
Open Source / Extensible	–	✓	✓
Algebraic calculations and aggregations	[limited]	✓	✓

http://github.com/modilabs/bamboo   http://bamboo.io   Modi Research Group, Columbia University