

How to restore faith in your data
Ask Katie Paine Measurement Hour
May 2014



Katie Delahaye Paine
CEO
Paine Publishing

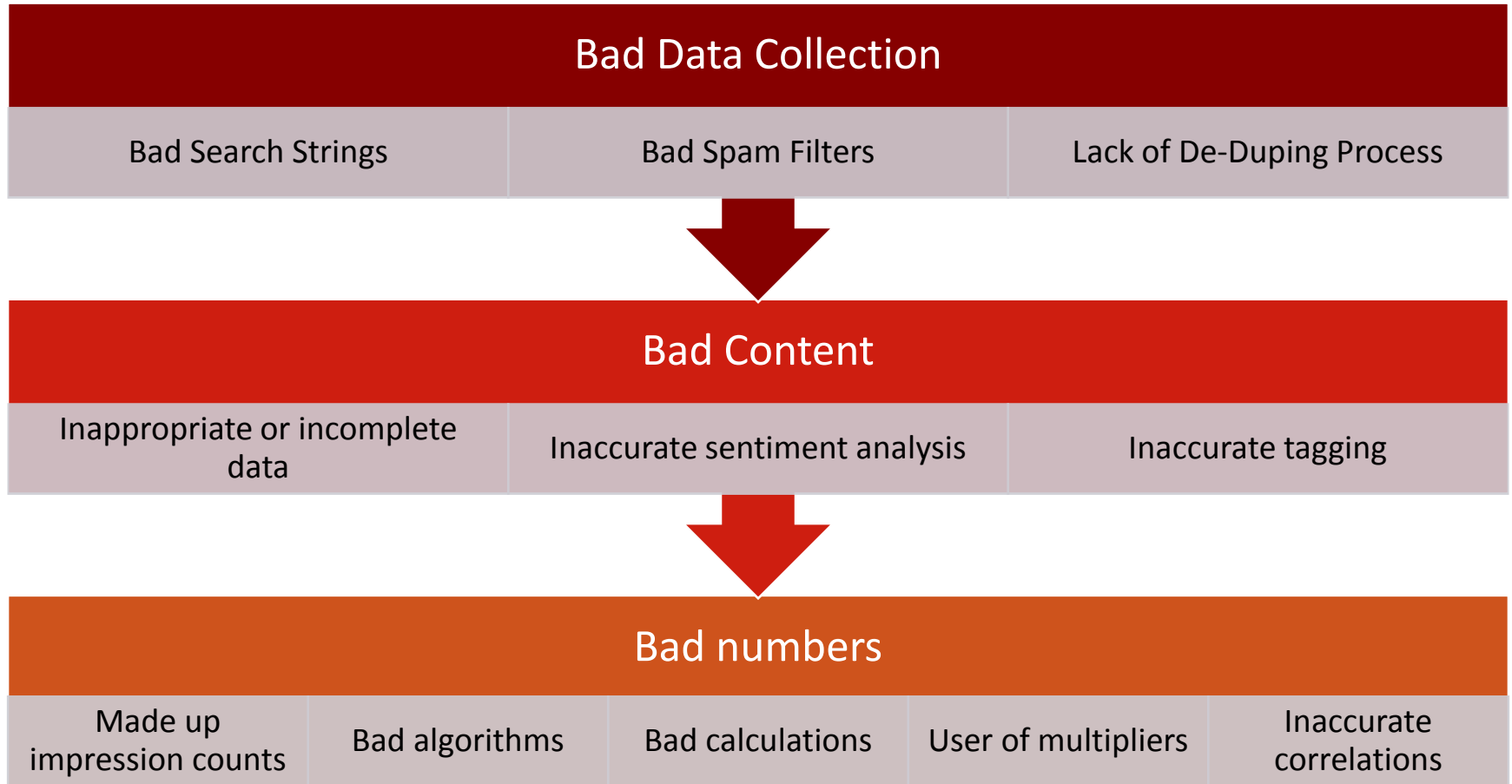
www.painepublishing.com
@queenofmetrics
measurementqueent@gmail.com

Steps to Data Integrity

- ▶ Collect good data
- ▶ Collect the right data
- ▶ Code it correctly
- ▶ Be conservative with your numbers
- ▶ Do the “So What” Test



The Bad Data Flow Chart



Where Bad Data Starts

Bad Data Collection

Bad
Search
Strings

Bad Spam
Filters

Lack of
De-Duping
Process

Step 1: Improve your Data Collection

▶ Bad Search Strings

- ▶ Need to be broad enough to capture the relevant items and narrow enough to eliminate spam
- ▶ Need to be specific
- ▶ Need to have terms a computer can understand
- ▶ Need to be up-to-date

▶ Bad Spam Filters

- ▶ 40% of what you bring in is probably irrelevant

▶ No De-Duping

- ▶ De-dupe on specific URL, Date, Outlet and Title

The Bad Data Flow Chart

Bad Content

Inappropriate or
incomplete data

Inaccurate
sentiment
analysis

Inaccurate tagging

Step 2: Improve your Content

- ▶ Is the Data In your Database What you want?
 - ▶ Inappropriate?
 - ▶ Incomplete?
- ▶ Is the coding accurate?
 - ▶ Does it conform to standards?
 - ▶ Is it consistent?
 - ▶ Does it follow an agreed upon set of coding instructions?
- ▶ Are items “bucketed” correctly
 - ▶ Tagging vs Coding
 - ▶ Why its important

The Bad Data Flow Chart

Bad numbers

Made up
impression
counts

Bad
algorithms

Bad
calculations

User of
multipliers

Inaccurate
correlations

Bad Numbers

- ▶ Made up impression counts
- ▶ Bad algorithms
- ▶ Bad calculations
- ▶ User of multipliers
- ▶ Inaccurate correlations

Thank You!

- ▶ For more information or copies of these slides go to www.painepublishing.com
- ▶ Email me at measurementqueen@gmail.com
- ▶ Follow me on Twitter: [@queenofmetrics](https://twitter.com/queenofmetrics)
- ▶ Friend me on Facebook: Katie Paine
- ▶ Or call me at 1-603-682-0735

