INDIANA UNIVERSITY

CENTER FOR RESEARCH AND EVALUATION
Lifelong Learning Group

Sense Making of Big Data

Front-end study on familiarity with terms

October, 2014

Prepared by:
Joe E. Heimlich, Ph.D
Zdanna Tranby
Mary Ann Wojton, Ph.D

Prepared for:
Sense-making of Big Data Team

This project was completed with support from the National Science Foundation (1223698)

LifelongLearningGroup.org
COSI | 333 West Broad Street | Columbus, Ohio 43215

Introduction

The *Sense Making of Big Data project* is designed to study how audiences in public spaces, in this case those in a museum setting, relate to and make sense of representations of large data sets. As the test case for complex big data representations, the project has used representations that provide an overview of science generally as well as specific areas of STEM, charting and exploring the history and future of science and technology. The data collection portion of the project has occurred at the New York Hall of Science, the Marian Koshland Science Museum in Washington D. C., COSI in Columbus, Ohio, and WonderLab Museum in Bloomington, Indiana. Findings from this project will inform the development of a traveling, hands-on exhibition that will enable visitors to create and utilize representations of big data such as maps and charts. In addition, the project hopes to create a foundation for the design of informal learning experiences that encourage participants to explore, engage and make better sense of big data. This project is potentially transformative because big data is becoming ubiquitous and making sense out of data representations is necessary in order to understand and begin to utilize big data.

Methods

To better help people make sense of large data sets, also called "big data", it would be invaluable to know the ways people understand (or not) the phrases "big data" and "data visualization." In moving forward, it could be important to know how people make meaning from these phrases.

The questions guiding this front-end study are:

- *Have people heard the phrase big data?*
- *What phrases do people use to explain it, if they have heard the phrase, or what they think it could be if they were not familiar with the phrase?*
- *Is there any familiarity with the phrase data visualization?*
- *Of these respondents, how do they identify their interests in STEAM?*

Population

The participants in this study were adult visitors to COSI (Columbus) and Science Museum of Minnesota (St. Paul) during late July and August, 2014. The goal was a maximum of 45 participants per museum.

Methodology

Using intercepts at both institutions, adult visitors were approached and invited to participate in a very brief interview. If the person agreed, they were asked additional questions. The first question asked about awareness of the phrase, "Big Data" or for a very small comparison group, "Data Visualization." Visitors were then asked "How would you explain "Big Data" to a friend?" if they had heard the phrase before. If they hadn't heard it, they were asked, "What does the phrase 'Big Data' make you think of?" Finally, they were asked to rate their interest in science, math, and art (scale of 1-10 from not-at-all interested to totally love it for each).

Results

A total of 92 interviews were conducted: 85 using the term Big Data and 7 using the term Data Visualization. There were at least 40 interviews conducted at each institution (SMM & COSI), 44 of the Big Data interviewees presented as male, and 48 presented as female.

Big Data

Awareness of the phrase

More than two-thirds of visitors interviewed said that they had not heard the phrase "Big Data" before (71%, n=85).

When asked how they would explain big data to a friend or what they thought the phrase made them think of, most people gave neutral explanations or associations with the phrase, but a handful had positive or negative reactions like these:

> "[It's] important information. Something everybody relates to, but doesn't understand."

> "[It] gives me anxiety. I don't know and I don't like it."

Most people, even those who had not heard the phrase 'Big Data' before, were able to make appropriate associations with the phrase (88%, n=85). However, those that said that they had heard of the phrase 'Big Data' were more likely to explain how it might be used to accomplish something (32%, n=25), than those who said that they hadn't heard of the phrase (7%, n=60) (see Table 1). Over 2/3 of those who had heard of 'Big Data' responded with complex, or definitional responses while slightly fewer than half of those who said they had not heard of 'Big Data' did so.

Table 1. Visitor associations with the phrase 'Big Data' (n=85)

Type of Response	Response Themes	Have heard of 'Big Data' (n=25)	Have NOT heard of 'Big Data' (n=60)
Complex	How Big Data can be used	32% (8)	7% (4)
Definition	Lots of data/information/numbers	36% (9)	40% (24)
Association	Tools for collecting or storing data	12% (3)	28% (17)
Loose Association	Other comments that might be related to Big Data	8% (2)	12% (7)
No Association	Did not talk about something related to Big Data	12% (3)	12% (7)

[If yes] "How would you explain "Big Data" to a friend?" (n=25)

Of those responding, eight (32%) described how Big Data can be used in their explanations:

- Some of these responses referred specifically to the large amounts of data, but also included elements of synthesis, meaning-making, or use:
    - "A lot of information. The industry and complex-ation of data, to provide scientific resources to compile information and conduct analysis for technology improvement."
    - "Large amount of information that is involved with brain process. To condense the information."
    - "Large amounts of data, categorized in a way you can do something meaningful with it."
    - "About trends. I know very little about it. Patterns of behavior, such as where there are infections, advantage of collecting on national level is that it can help prevent spread of infection. Exciting concept."

- Other comments referred to the collection of information and, in many cases, the collection for commercial purposes:
    - "Collecting information that is helpful for people to make conclusions."
    - "The massive collection of information collected by internet companies on us; that can be mined and sold."
    - "The private and public sectors collect information, the accumulating volume of information that is hard to comprehend. The fast accumulation of all the data in our routine transactions."
    - "A digital footprint. All the tracking of anything electronic, for example, what you purchase, credit scores."

An additional 9, or 36%, of these respondents referred to amounts of data, information, or numbers:

- Several of the comments referred to amounts of information:
    - "Giant databases. Works in storage with tera-data and computational datasets."
    - "Gigantic data projects."
    - "Huge amount of data."
    - "Large data set, lot of proof for something."
    - "Lots of compiled data."
    - "Lot of information that lives in a computer."
    - "Amount of data, volume, velocity and variety (audio, text, etc)."

- A few referred to collection of amounts of data:
    - "Collecting lots and lots of data."
    - "Collection of information. A pile of information on the internet, such as government documents and company files."

Three individuals saw Big Data as tools for collecting or storing data:

- "Don't know. Cloud."
- "IBM is doing something with it. IT stuff, technology."
- "Storage, computer-wise, where we store our photos in."

Two people offered other comments that might be related to Big Data:

- "Collect interactions with a statistic plan."

- "Something to do with technology and informatics."

And three people, in their explanations, did not talk about something related to Big Data:

- "A lot of objects in the mall."
- "Don't know. Perhaps I wouldn't explain it to a friend."
- "No idea."

[If no] "What does the phrase "Big Data" make you think of?" (n=60)

For this group of respondents, only 4 or 7% of the respondents included statements related to how Big Data can be used:

- "A bunch of data. Data that reveals big results, significant results."
- "Digital life. To analyze customer behaviors."
- "Don't know. Important information, something everybody relates to, but doesn't understand."
- "Researchers collect a lot of data and try to make sense of it."
- "Data mining."

Twenty-four, or 40% of the respondents' comments referred to lots of data, information, or numbers.

- Some of these comments specifically tied to a data set of computer-based information/files:
    - "A huge data set with significant sample size."
    - "A lot of systematic information in computers."
    - "A large computer file."
    - "A large database. Or something that contains a lot of data."
    - "A lot of data or a really large computer."
    - "Big database of consumer information."

- Many of these comments simply reiterated the "big" in referring to lots of data:
    - "A lot of information, facts, and figures, statistics."
    - "A lot of numbers."
    - "Extreme. Lots of data."
    - "Huge amount of data."
    - "Massive amounts of data."
    - "Information."
    - "Information. Meta-information, not minutia."
    - "Large data."
    - "Large numbers in relationship to a small concept."
    - "Lots of information. A company that has a big database."
    - "Lots of information; overwhelming amounts of information."
    - "Lots of number collecting."
    - "Maybe data from… No idea. Makes me think of an experiment. Maybe data on really large scale."
    - "Gathering lots of info on a certain subject."
    - "Numbers."

Slightly more than a quarter of these respondents (17 or 28%) suggested tools for collecting or storing data:

- Technology.
    - “High tech, age of technology. Getting bigger and bigger.”
    - “Cloud. “
    - “Lots of technology.”
    - “More cell phones, iPads, collection of data”
- Several referred to computers and data in computers
    - “Computer programs.”
    - “Computers.”
    - “Computers. So much data in computers.”
- Data collection, processing, storage, etc. I work in IT.
    - “Surveys.”
    - “Something on the computer, collection of stuff.”
- Star Trek episodes, the character "Data." A bunch of computers.

Other comments that might be related to Big Data were offered by 7 (12%) of these respondents:

- “Computer information.”
- “Data entry.”
- “How we collect data to decide if the data is valid or not.”
- “Meta-data. Collecting meta-data on us, like the NSA.”
- “NSA.”
- “Research.”
- “Some tech piece and a large study.”

One eighth of the respondents (7 or 12%) did not talk about something related to Big Data

- “Don't know about it.”
- “Like a punctuation mark.”
- “Monster.”
- “No idea.”

Data Visualization

Seven additional interviews were conducted at COSI where the phrase “Data Visualization” was used in place of the phrase “Big Data” in the questions above. Two of the seven said that they had heard the phrase. They said that they would explain it to a friend like this:

- “[A] way to see information in way that is useful.”
- “Graphs and charts that make [sense of] something to look at.”

Those who hadn’t heard of that phrase before said that it made them think of:

- “Numbers and stuff. Information.”
- “Not even sure.”
- “Innovation.”
- “A way to see data.”
- “Memorizing data.”

We chose not to pursue a large sample size investigating data visualizations because this team has already conducted several studies into visitor understanding of big data visualizations and how they make sense of them.

Interest in Science, Math, and Art

Continuing a question asked in other front-end studies, we asked about visitor interest in Science, Math, and Art. Both science and art had close, very strong scores with a mean of 7.9 for science and 8.0 for math on a ten-point scale. Math was much lower with a mean of 6.6. Both science and art also had modes (in yellow in the table) of 10 and medians of eight (blue). Math had both mode and median at 7.

Table 2. Visitor interest in Science, Math, and Art

	Science (n=92)	Math (n=91*)	Art (n=92)
1 (Not at all interested)	1	6	-
2	-	4	-
3	1	2	1
4	4	3	2
5	5	13	9
6	6	6	9
7	13	16	14
8	24	16	14
9	11	9	19
10 (Totally love it)	27	16	24
Average Score	8.0	6.7	7.9

*All 85 interviewees were asked to rank their interest in Math, but one responded "0", and is not included in this table

Conclusions and Recommendations

The majority of respondents did not have top of mind recall of the phrase "big data." Even so, a plurality of both those who did have recall and those who did not were able to suggest that the phrase refers to amounts of data, thought the circularity of that definition might suggest a non-critical understanding. Almost a third of those who did have top of mind recall described big data by referring to use of data.

Compared to the respondents to the phrase "big data," there was a very similar proportion of awareness of the phrase "data visualization" observed in the small, comparative study respondents.

Although the phrase "big data" is ubiquitous, it appears it may not carry the meaning among visitors to the science centers as it does to those who use the phrase with an intentional meaning.

There is a group of respondents who see big data from the marketing/web-based data use, and may have a bias toward big data from this focused perception. The value and use of big data beyond the marketing focus may need to be emphasized to facilitate understanding of the breadth of use.