NSA Collection of “Metadata”

● What is this report saying?
● According to this report, did the NSA listen to our actual phone conversations?
● Which information then did they collect?

http://mashable.com/2013/06/07/nsa-prism-faq/
**Metadata**

- Data Dictionary
- “Data for data”
- Data that explains what this data is, what it is for, and how it is structured
  - e.g. title, description, data type
The ‘Me’ in Metadata

Almost every digital file we generate carries invisible tags.

**METADATA EXAMPLES**
- Focal length
- Camera type
- Date & time taken
- Exposure
- Flash setting
- Internet provider
- Mail client
- IP address
- Preferred language
- Home location
- Place ID

From the tweet:
- COORDINATES: 42.59640, -114.4032
- LANGUAGE: EN

From the email:
- DATE/TIME: 2013: 04: 22 15: 57: 33

From the photo:
- CAMERA TYPE: HTC ONE X
- FLASH SETTING: AUTO

1. Geoff poses by a waterfall and snaps a self-portrait, which he immediately tweets, then emails to his grandmother.
2. Geoff’s text, photo, and email ascend to a series of remote servers, each dragging their own trails of metadata.
3. Once there, the metadata may be extracted and interpreted by any interested party with access.

Thanks to the above metadata, without ever having met Geoff, we know he was at Shoshone Falls near Twin Falls, Idaho, at 3:57 p.m. on April 22, that he has an HTC One X smartphone, and that he is an English language speaker.

Note: Actual metadata code modified for readability. Many metadata values, such as time, can be extracted from multiple sources, and values may differ slightly.

Source: staff reporting

The Wall Street Journal

## How would you describe this data?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>DC_KEY</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>OBJECTID</th>
<th>TEXT_GENERAL_CODE</th>
<th>POINT_X</th>
<th>POINT_Y</th>
<th>SHAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>14</td>
<td>N</td>
<td>2010-05-05 11:34</td>
<td>2010-05-05 11:34:00</td>
<td>1.99814E+11</td>
<td>300 BLOCK E CLIVEDEN</td>
<td>300</td>
<td>12074596 Robbery No Firearm</td>
<td>-75.1772</td>
<td>40.05271</td>
<td>12028557</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>A</td>
<td>2010-03-31 12:47</td>
<td>2010-03-31 12:47:00</td>
<td>1.99819E+11</td>
<td>6000 BLOCK HAVENFORD</td>
<td>600</td>
<td>1207790 Thefts</td>
<td>-75.2402</td>
<td>39.9705</td>
<td>12032151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>1</td>
<td>2014-09-14 16:00</td>
<td>2014-09-14 16:00:00</td>
<td>1.99822E+11</td>
<td>1800 BLOCK W MONTGOMERY</td>
<td>100</td>
<td>12091174 Homicide - Criminal</td>
<td>-75.1568</td>
<td>39.98804</td>
<td>12046078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>L</td>
<td>2007-08-13 7:40</td>
<td>2007-08-13 7:40:00</td>
<td>2.00019E+11</td>
<td>0 BLOCK N 62ND ST</td>
<td>600</td>
<td>12075475 Theft from Vehicle</td>
<td>-75.2448</td>
<td>39.96251</td>
<td>12029336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>E</td>
<td>2011-08-04 12:49</td>
<td>2011-08-04 12:49:00</td>
<td>2.00038E+11</td>
<td>3600 BLOCK RED LION RD</td>
<td>600</td>
<td>12228454 Thefts</td>
<td>-74.9862</td>
<td>40.07513</td>
<td>12182815</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>U</td>
<td>2011-09-19 6:16</td>
<td>2011-09-19 6:16:00</td>
<td>2.00039E+11</td>
<td>5000 BLOCK GERMANY</td>
<td>600</td>
<td>12201847 Thefts</td>
<td>-75.1655</td>
<td>40.03021</td>
<td>12156208</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>N</td>
<td>2011-08-03 12:43</td>
<td>2011-08-03 12:43:00</td>
<td>2.00402E+11</td>
<td>1000 BLOCK ROSALIE ST</td>
<td>600</td>
<td>12076028 Theft from Vehicle</td>
<td>-75.0897</td>
<td>40.03553</td>
<td>12038589</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>M</td>
<td>2007-09-10 7:32</td>
<td>2007-09-10 7:32:00</td>
<td>2.00415E+11</td>
<td>6500 BLOCK MARSDEN</td>
<td>700</td>
<td>12096073 Recovered Stolen Motor Vehicle</td>
<td>-75.0492</td>
<td>40.02361</td>
<td>12045034</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>I</td>
<td>2011-08-05 12:57</td>
<td>2011-08-05 12:57:00</td>
<td>2.00415E+11</td>
<td>2100 BLOCK BETTS ST</td>
<td>600</td>
<td>12201848 Thefts</td>
<td>-75.0682</td>
<td>40.01318</td>
<td>12156209</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>35</td>
<td>H</td>
<td>2008-04-14 8:38</td>
<td>2008-04-14 8:38:00</td>
<td>2.00435E+11</td>
<td>900 BLOCK W LOUDEN ST</td>
<td>700</td>
<td>11068771 Recovered Stolen Motor Vehicle</td>
<td>-75.1391</td>
<td>40.02502</td>
<td>11561132</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>39</td>
<td>S</td>
<td>2007-08-10 11:18</td>
<td>2007-08-10 11:18:00</td>
<td>2.00539E+11</td>
<td>5000 BLOCK COPLEYS RD</td>
<td>700</td>
<td>12070695 Motor Vehicle Theft</td>
<td>-75.1742</td>
<td>40.02045</td>
<td>12034556</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>G</td>
<td>2006-01-01 14:08</td>
<td>2006-01-01 14:08:00</td>
<td>2.00601E+11</td>
<td>1500 BLOCK W MOYAM</td>
<td>600</td>
<td>12072181 Thefts</td>
<td>-75.1751</td>
<td>39.91645</td>
<td>12026542</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>J</td>
<td>2006-01-01 17:26</td>
<td>2006-01-01 17:26:00</td>
<td>2.00601E+11</td>
<td>2400 BLOCK S 24TH ST</td>
<td>600</td>
<td>12070915 Thefts</td>
<td>-75.187</td>
<td>39.92192</td>
<td>12033757</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>J</td>
<td>2006-01-02 10:12</td>
<td>2006-01-02 10:12:00</td>
<td>2.00601E+11</td>
<td>S 23RD ST /RTNER ST</td>
<td>600</td>
<td>12069241 Thefts</td>
<td>-75.1852</td>
<td>39.92258</td>
<td>12024114</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>A</td>
<td>2006-01-02 12:36</td>
<td>2006-01-02 12:36:00</td>
<td>2.00601E+11</td>
<td>1400 BLOCK M IFLIN ST</td>
<td>600</td>
<td>12067807 Theft from Vehicle</td>
<td>-75.1693</td>
<td>39.92678</td>
<td>12022168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>F</td>
<td>2006-01-03 7:05</td>
<td>2006-01-03 7:05:00</td>
<td>2.00601E+11</td>
<td>2200 BLOCK S 20TH ST</td>
<td>600</td>
<td>12074140 Thefts</td>
<td>-75.1797</td>
<td>39.92335</td>
<td>12028491</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>D</td>
<td>2006-01-03 12:31</td>
<td>2006-01-03 12:31:00</td>
<td>2.00601E+11</td>
<td>S 29TH ST /MC KEAN ST</td>
<td>300</td>
<td>12165554 Robbery No Firearm</td>
<td>-75.1939</td>
<td>39.92874</td>
<td>12119151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>C</td>
<td>2006-01-03 14:51</td>
<td>2006-01-03 14:51:00</td>
<td>2.00601E+11</td>
<td>1700 BLOCK W MCGOWAN</td>
<td>300</td>
<td>12066171 Robbery No Firearm</td>
<td>-75.1473</td>
<td>40.01323</td>
<td>12045323</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metadata for Philadelphia Crime Logs

**Key attribute field names and descriptions:**

- **DC_Dist:** a two character field that identifies the District boundary.
- **Sector:** a single character field that identifies either the newer Police Service Area (PSA) boundary or the obsolete Sector boundary. This field was converted from Sectors to Police Service Areas in 2009 as a new crime initiative.
- **Dispatch_Date/Dispatch_Time:** date and time that the officer was dispatched to the crime scene.
- **Hour:** generalized hour of the dispatched time.
- **DC_Key:** unique identifier of the crime that consists of Year + District + Unique ID.
- **Location_Block:** location of crime generalized by street block.
- **UCR_General:** rounded crime code, i.e. 614 to 600.
- **Text_General_Code:** generalized text/description of the crime code.
- **Point_X:** Longitude of crime location.
- **Point_Y:** Latitude of crime location.
# Data Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Examples</th>
<th>What it’s called in Excel</th>
<th>What it’s called in Tableau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integer</td>
<td>Whole numbers</td>
<td>35, 102, -40, 0</td>
<td>Number</td>
<td>Number (whole)</td>
</tr>
<tr>
<td>Floating point</td>
<td>Fractional values</td>
<td>3.56, -1.0, 10.123</td>
<td>Number</td>
<td>Number (decimal)</td>
</tr>
<tr>
<td>Boolean</td>
<td>Binary (2) values</td>
<td>True/False, Male/Female</td>
<td>N/A</td>
<td>Boolean</td>
</tr>
<tr>
<td>String</td>
<td>Numeric and non-numeric characters</td>
<td>Bob, I like cheese, hello123</td>
<td>Text</td>
<td>String</td>
</tr>
<tr>
<td>Date/Time</td>
<td>Calendar date and time</td>
<td>8/31/2014, 10:05 AM, 8/31/2014 10:05 AM</td>
<td>Date</td>
<td>Date or Date &amp; time</td>
</tr>
</tbody>
</table>
Why do we need Metadata?

- Because computers are dumb!
- It is getting there, but computers are still not able to understand what the data means and what it is for.
- So, we have to describe it for them.

http://www.manning.com/sande/excerpt_pREFace.html
Metadata is everywhere

How Google Works [Kindle Edition]
Eric Schmidt (Author), Jonathan Rosenberg (Author)

Print List Price: $30.00
Kindle Price: $10.99
You Save: $19.01 (63%)
Sold by: Hachette Book Group

- Word Wise: Enabled
- Don't have a Kindle? Get your Kindle here.
- Whispersync for Voice: Ready

Free Kindle Reading App
Anybody can read Kindle books—even without a Kindle device.

The New York Times
http://nyti.ms/1sSSAQD

N.Y. / REGION

Cuomo to Propose 2 New Minimum Wages, for New York City and State

By JESSE McKINLEY JAN. 18, 2015