Extensive datasets are not today’s phenomenon

<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1959</td>
<td>Data set of 1,000 cases</td>
</tr>
<tr>
<td>1965</td>
<td>Data sets of 10,000 cases (e.g. observations)</td>
</tr>
<tr>
<td>1978</td>
<td>Technology to analyse 500 to 5,000 variables at the same time</td>
</tr>
<tr>
<td>1981</td>
<td>Statistics for 20,000 samples, as histograms with 800 cells</td>
</tr>
<tr>
<td>1981</td>
<td>Software to handle 88,000 variables</td>
</tr>
<tr>
<td>1982</td>
<td>Moderate data sets have less than 500 cases and large more than 2,000</td>
</tr>
<tr>
<td>1986</td>
<td>Very large data sets referred as 11,000 cases</td>
</tr>
<tr>
<td>1987</td>
<td>50,000 points, but the representation of 1 million or more data is feasible</td>
</tr>
<tr>
<td>1990</td>
<td>5,000 cases with 6 variables for immediate evaluation (c. 3 seconds.)</td>
</tr>
<tr>
<td>1991</td>
<td>1 million cases a very large data file</td>
</tr>
<tr>
<td>1996</td>
<td>Large data sets have 60,000 or more observations</td>
</tr>
<tr>
<td>2006</td>
<td>1 million (bytes, cases, variables, combinations or tests) as a guideline to refer to a dataset as ‘large’</td>
</tr>
</tbody>
</table>
The projected volume of ESA's EO Data Archives in 2020 (estimated in 2013) will exceed 2.4 PB. The first three Sentinel satellites alone are generating products having as much data inside as 3 PB/year, a size which will be doubled when their twins will be in full operations.
Volume of Our Big Data

ASI COSMO-SkyMed

Today 0.4 PB

Up to 0.6 PB in 2020
Veracity of Big Data

Data quality

Data accuracy

The concept of “fit for purpose”
Big Data
Improve the ability to analyze and shape the real world.
Big Data enable decision making process.
Value of Big Data

Earth Observation

Business Innovation

Free, Full and Open Data Policies
Techniques of Big Data

SENSORS > Smaller, smarter and more heterogeneous

TELECOMM > More capacity

COMPUTING > Supercomputer For Everyone

STORAGE > 5.5 PB on one cubic millimeter of DNA molecules

INTERNET CONNECTIVITY > 43% of world population

CLOUD COMPUTING > Anytime, anywhere, access to IT

DATA MINING TECHNIQUES > Advanced queries