Pattern recognition by humans and machines over large data sets

C. Versino

European Commission
Joint Research Centre (JRC)
Institute for Transuranium Elements (ITU)
Nuclear Security Unit
Ispra, Italy
Outline
‘Data retrieval and analysis over large data sets’

Issues
Will present main **issues** in data retrieval/analysis,

Technology
and highlight ways of using information technology, based on data visualisation, to address these issues.

Examples
Will present **example visualisations** related to nuclear safeguards.

- Invisible Big Data
- Data access
- Precision vs Accuracy of information

CN 220-224 Tools for video reviews
CN 220-293 Tools for trade analysis...
Invisible Big Data

Large data sets are buried in databases and repositories. We do not see data like we see the world around us. There is a narrow communication channel between the data and the user (even if you are feeling lucky).
Issue

Data access

In many cases data access is mediated by queries. One needs to formulate useful queries before seeing any data. Only slices of filtered data are returned. Little data integrity.

By contrast a data visualisation approach would feature the data first. Seeing the data distribution may trigger questions that one would not have imagined otherwise.
“Even if the amount of knowledge in the world is increasing, the gap between what we know and what we think we know may be widening. This syndrome is often associated with very precise-seeming predictions that are not at all accurate. (...) This is like claiming you are a good shot because your bullets always end up in about the same place — even though they are nowhere near the target.”

Nate Silver
The Signal and the Noise
The data visualisation process

**Effort**

<table>
<thead>
<tr>
<th>5%</th>
<th>90% of time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>20% of time</td>
</tr>
<tr>
<td>5%</td>
<td>30% of time</td>
</tr>
</tbody>
</table>

**Data gathering**
Queries on third parties DBs, sensor data, own generated data, ...

**Data preparation for analysis (analysis with IT)**
Data de-structuring to raw format, + meta-data

**Data visualisation**
Encode abstract data in graphical form for analysis and communication.

- Explore
- Understand
- Question ...
- Make a point
- Findings
- Report ...

**Analytical interactions:** adding / removing dimensions, sorting, filtering, highlighting, aggregating / disaggregating, drilling, grouping, zooming/panning, re-visualising, re-expressing, re-scaling ...

Enables human visual recognition.
Works pre-attentively.
Parallel (high bandwidth).

→ Fast
Technology

Raw data – Data integrity – Data sushi

Data sushi:

‘A visualisation which is beautiful on the outside and has raw data on the inside’

Jock Mackinlay
Jock’s Dream of Data Sushi

Why using raw data is important?

• Gives the analyst the ability to create **overviews** of the data (data integrity, accuracy, completeness) and **detailed views** as required (precision, correctness).

• Result data views are generated on demand as visual cross-tabs of data dimensions of interest to the analyst (i.e., not decided by a data provider as pre-defined views or paths to get to the data).

• ‘Validates the author’ of data views (peers can explore the same data set and confirm or find different/other/more results).

• Facilitates blending of other data sources (adding more dimensions, relate with independent sources).

• …
Example

Safeguards video reviews

Data visualisation – Overview first

S. Blunsden, C. Versino
VideoZoom storyboard
Example

Data visualisation – Details on demand

Safeguards video reviews

S. Blunsden, C. Versino
VideoZoom zooming interface
Example

Data visualisation – Raw data
Example

Data visualisation – Data composition

Nuclear trade analysis
Import Export databases
Country X
Nuclear Trade Profile

Import Export Balance
2012-2014
29%
71%

Import share
Top import from
United Kingdom
Russia
Netherlands
Germany
Canada
Australia
France
Kazakhstan
Namibia
China
Sweden
Japan

Export share
Top export to
Japan
Taiwan
Canada
Korea, South
United Kingdom
China
Sweden
Russia
Mexico
Germany
Spain
France

Commodity import

Commodity export

Example
Nuclear trade analysis
Import Export databases
Example

Data visualisation – Details on demand

Country X

Nuclear Trade Profile

### Import 2014

<table>
<thead>
<tr>
<th>HS6 CODE</th>
<th>HS6 DESCRIPTION</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
</tr>
</thead>
<tbody>
<tr>
<td>2612.10</td>
<td>Uranium ores and concentrates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.10</td>
<td>Natural uranium and its compounds; alloys, dispe.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.20</td>
<td>Uranium enriched in U 235 and its compounds: pl.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.30</td>
<td>Uranium depleted in U 235 and its compounds, th.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.40</td>
<td>Radioactive elements, isotopes and compounds, th.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.50</td>
<td>Spent “irradiated” fuel elements “cartridges” of nu.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2845.10</td>
<td>Heavy water “deuterium oxide” [Euratom]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2845.90</td>
<td>Non-radioactive isotopes, inorganic or organic co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8109.90</td>
<td>Articles of zirconium, n.e.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.10</td>
<td>Nuclear reactors [Euratom]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.20</td>
<td>Machinery and apparatus for isotopic separation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.30</td>
<td>Fuel elements “cartridges”, non-irradiated, in casi.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.40</td>
<td>Parts of nuclear reactors, n.e.s. [Euratom]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Export 2014

<table>
<thead>
<tr>
<th>HS6 CODE</th>
<th>HS6 DESCRIPTION</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
</tr>
</thead>
<tbody>
<tr>
<td>2612.10</td>
<td>Uranium ores and concentrates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.10</td>
<td>Natural uranium and its compounds; alloys, dispe.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.20</td>
<td>Uranium enriched in U 235 and its compounds: pl.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.30</td>
<td>Uranium depleted in U 235 and its compounds, th.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.40</td>
<td>Radioactive elements, isotopes and compounds, th.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2844.50</td>
<td>Spent “irradiated” fuel elements “cartridges” of nu.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2845.10</td>
<td>Heavy water “deuterium oxide” [Euratom]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2845.90</td>
<td>Non-radioactive isotopes, inorganic or organic co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8109.90</td>
<td>Articles of zirconium, n.e.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.10</td>
<td>Nuclear reactors [Euratom]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.20</td>
<td>Machinery and apparatus for isotopic separation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.30</td>
<td>Fuel elements “cartridges”, non-irradiated, in casi.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8401.40</td>
<td>Parts of nuclear reactors, n.e.s. [Euratom]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example

Data visualisation – Details on demand

Country X
Nuclear Trade Profile

<table>
<thead>
<tr>
<th>HS6 CODE</th>
<th>HS6 DESCRIPTION</th>
<th>PARTNER</th>
<th>VALUE USD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2845.10</td>
<td>Heavy water</td>
<td>Australia</td>
<td>$0.00M</td>
</tr>
<tr>
<td></td>
<td>“deutornium oxide”</td>
<td>Canada</td>
<td>$2.00M</td>
</tr>
<tr>
<td></td>
<td>[Euratom]</td>
<td>China</td>
<td>$4.00M</td>
</tr>
</tbody>
</table>

FLOW / PERIOD

<table>
<thead>
<tr>
<th></th>
<th>Import 2014</th>
<th>Export 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jan Feb Mar</td>
<td>Jan Feb Mar</td>
</tr>
</tbody>
</table>

- Germany
- India
- Japan
- Malaysia
- Mexico
- Netherlands
- Nigeria
- Qatar
- Switzerland
- Taiwan
- United Arab Emirates
- United Kingdom
Conclusions

• Issues in data retrieval and analysis arise when:
  • The data are ‘invisible’
  • Data access starts by questions and not by data presentation
  • Retrieval and analysis systems strive more for results’ precision (correctness) than accuracy (completeness).

• Data visualisation approaches can mitigate these issues in that priority is given to data presentation.
  This encourages data exploration by the analyst, enabling more accurate results and higher data integrity.

• A key point, often not understood, is that data visualisation requires working with raw data, not ‘result set data’.
Acknowledgements

The work presented is funded by the European Commission, Joint Research Centre, in projects: VideoZoom and Strategic Trade Analysis for Non Proliferation. Both projects contribute to the EC Support to the IAEA.

References


https://www.youtube.com/watch?v=EsyMkuMM8HU


