iSHARE2 Core Micro Dataset Quality

Tathagata Bhattacharjee

iSHARE2 Support Team
Outline

- Understanding the concept of Data Quality
- Data Quality Rules for Core Micro Dataset

Session duration: 15 minutes
Data Quality

• Data are of high quality "if they are fit for their intended uses in operations, decision making and planning" (J. M. Juran, evangelist for quality)

• The measurement of quality of data can be expressed by its completeness, validity, consistency, timeliness and accuracy that will make the data appropriate for a specific use

• The processes and technologies involved in ensuring the conformance of data values to a specific requirements and acceptance criteria is quality assurance
Quality Checks done in iSHARE2 ETL Process

• Ensuring the first event is legal
  – Like the first event must be **Enumeration, Birth** or **Inmigration**.

• Ensuring the last event is legal
  – Like the last event must be **End of Observation, Death** or **Outmigration**.

• Ensuring the event transitions are legal
## Transition Checks

<table>
<thead>
<tr>
<th>After Event (Followed By)</th>
<th>ENU</th>
<th>BTH</th>
<th>IMG</th>
<th>EXT</th>
<th>ENT</th>
<th>DTH</th>
<th>OMG</th>
<th>OBE</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENU</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>GREEN</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>BTH</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>IMG</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>EXT</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>ENT</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>DTH</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>OMG</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>OBE</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
</tbody>
</table>

- **Legal Transition**
- **Illegal Transition**
- **Legal in case of multiple movements**

iSHARE 2  
an INDEPTH project
Quality Metric

• Quality Metric 1: % of all start events that are illegal
• Quality Metric 2: % of all end events that are illegal
• Quality Metric 3: % of all transitions that are illegal
Thank you