**NBD(NIST Big Data) Requirements WG Use Case Template**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | Big Data Archival: Census 2010 and 2000 – Title 13 Big Data | |
| **Vertical (area)** | | Digital Archives | |
| **Author/Company/Email** | | Vivek Navale & Quyen Nguyen (NARA) | |
| **Actors/Stakeholders and their roles and responsibilities** | | NARA’s Archivists  Public users (after 75 years) | |
| **Goals** | | Preserve data for a long term in order to provide access and perform analytics after 75 years. | |
| **Use Case Description** | | 1. Maintain data “as-is”. No access and no data analytics for 75 years. 2. Preserve the data at the bit-level. 3. Perform curation, which includes format transformation if necessary. 4. Provide access and analytics after nearly 75 years. | |
| **Current**  **Solutions** | **Compute(System)** | | Linux servers |
| **Storage** | | NetApps, Magnetic tapes. |
| **Networking** | |  |
| **Software** | |  |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Centralized storage. |
| **Volume (size)** | | **380 Terabytes.** |
| **Velocity**  **(e.g. real time)** | | **Static.** |
| **Variety**  **(multiple datasets, mashup)** | | **Scanned documents** |
| **Variability (rate of change)** | | **None** |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues)** | | **Cannot tolerate data loss.** |
| **Visualization** | | **TBD** |
| **Data Quality** | | **Unknown.** |
| **Data Types** | | **Scanned documents** |
| **Data Analytics** | | **Only after 75 years.** |
| **Big Data Specific Challenges (Gaps)** | | Preserve data for a long time scale. | |
| **Big Data Specific Challenges in Mobility** | | TBD | |
| **Security & Privacy**  **Requirements** | | Title 13 data. | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | | . | |
| **More Information (URLs)** | |  | |