The Data Quality Lifecycle

Peter Hanson, MRICS
Senior Project Manager
Geospatial Data Services
Infotech Enterprises Europe Ltd

Data Quality Lifecycle

- Data Quality Issues
- Data Quality Measurement
- Applying PPM Principles
- Achieving PPM Goals

What is the Data Quality Lifecycle?

Infotech
Creating Business Impact

- Typically two stages:
  - Initial capture or conversion
  - Ongoing maintenance, including migration
- Process often takes into account multiple data sets of various degrees of known quality
- Need to understand the quality of the input data sources
  - Source Record Analysis
  - Data profiling
  - Establish priority of sources
- Prevent erosion of quality over time

Data Quality Issues

- Often too much focus on system functionality rather than data quality
- Data Quality Management not always top priority for system functionality
- Data Quality starts with the data capture (aka collection, conversion, acquisition)
- Legacy data can also be an issue –
  - Often seen as a minor inconvenience associated with a system replacement
  - Too little time dedicated to the transfer of massive amounts of data.
  - Little or no impact assessment – Systems / Business Efficiency

How Can Data Quality be Measured?

Infotech
Creating Business Impact

- **Topological Accuracy (start, not an end in itself)**
- **Four Key Elements:**
  - **Accuracy**
  - **Concurrency**
  - **Completeness**
  - **Spatial Integrity**
- **True data quality can only be measured when all are considered**

Accuracy – what does this mean?

www.infotechsw.com

Infotech
Creating Business Impact

- Issues related to GIS data quality:
  - Connectivity
  - Associations
- Specific applications require sophisticated data
- Spatial and A-Spatial / Attribution component.
- Fall between the classical data checks
- Crucial to the quality and usability of the data.
  - It's great to have all of one attribute 100% accurate but if it has no impact on the business, being more of a cosmetic attribute of the data than the quality
  - Each object/attribute has a data quality priority
  - More than the optional/mandatory flags
  - Gives focus on the key data areas.

Concurrency ? Completeness ?

Infotech
Creating Business Impact

- **Concurrency**
  - Increasingly becoming an issue
  - Organisations looking for 'real-time' views of GI data
  - Critical measure of quality
  - Complex – data needs to be time-stamped
  - Data may be pre-recorded to show the potential existence of data.
- **Completeness**
  - Not so easy to measure, especially during early stages of capture
  - Processes set up for tracking changes to data
  - Feedback loop on the data quality / completeness

Spatial Integrity

- **Just heard from 1-Spatial on this – need I say more??**

Applying PPM Principles to the Data Quality Lifecycle

Specification

- Changes to specifications should be discussed, agreed and maintained.
- Employ methods for ensuring currency of versions
- Set realistic AQLs
- Understand what AQLs actually mean with regards to the data

Improve Workflow

Infotech
Creating Business Impact

- Use WFM techniques
  - Systematic Approach
  - Rules-base driven
  - Embedded quality standards
  - In-built Root Cause Analysis
  - Process models
- Replication of customer environment
- Supplier involvement during data load – establish 'Quality Gates'
- Customer representatives at supplier site
- Ensure Production Life-cycle
- Continual monitoring
  - Process
  - People
- Query Management System

Quality

- Checks built-in to processes
- Effective control systems
- Proven QA techniques
- Double-entry methodology

Achieving PPM Goals

Procurement

Customers

- Share as much background information to the project
- Share information on future plans for data – context is very useful
- Seek help of suppliers who see things from a capture perspective
- Fully understand what "Fitness for Purpose" is

Suppliers

- Share experiences of similar work
- Make recommendations at pre-tender stage
- Share benefits of improvements in production processes

Evaluating the Supplier

- Approach to partnership
- Track record (especially where projects have not been successful)
- Accuracy of marketing
- Domain expertise of key staff
- Relevant experience
- Levels of staff attrition
- Attitude to risk management
  - On-going
  - Logs
  - Meetings
- Corporate Governance

Conclusions

www.infotechsw.com

Infotech
Creating Business Impact

- No silver bullet to solving data quality
- Organisations need to look at all aspects of the GIS deployment.
- Changes to the way systems are developed, deployed and maintained
- Quality is not just something that needs to be measured on the initial creation or from time to time
- Data quality and the continual improvement of quality, needs to be part of every interaction with the data
  - Applications
  - Processes
  - People.
- Data and quality are explicitly linked in the Data Life Cycle

Questions?