

# Applying Data Quality Standards to National Gazetteers

The NLPG – A Case Study

[rduffield@intelligent-addressing.co.uk](mailto:rduffield@intelligent-addressing.co.uk)

# The NLPG



## BLPU

- UPRN
- Easting & Northing
- Classification
- Organisation

## LPI

- LPI\_KEY
- UPRN
- USRN
- Language
- Secondary Addressable Object
- Primary Addressable Object
- Posttown & Postcode

## Street

- USRN
- Street Type
- Easting & Northings

## Street Descriptor

- USRN
- Street Descriptor
- Locality
- Town
- Administrative Area
- Language

## BLPU

- 10010224805
- (539131, 180124)
- Parent Shell Property Shell
- “AEG EUROPE”

## LPI

- 5330L000124603
- 10010224805
- 20803001
- English
- SAO = blank
- PAO = “The 02”
- Non-postally addressable

## Street

- 20803001
- Officially Named
- (539185, 179841) to (539200, 180027)

## Street Descriptor

- 20803001
- “Peninsula Way”
- Locality = blank
- “Greenwich”
- “London Borough of Greenwich”
- English

# The 02, Peninsula Square, Greenwich

# Components of data quality management



Governance



Standardisation



Continual improvement



Gathering Requirements



Measurement



Monitoring and reporting

# The ISO 19000 family

- Suite of inter-related quality standards for geographic information:
  - **ISO 19113** on concepts,
  - **ISO 19114** on principles for quality evaluation
  - **ISO 19138** quality assessment
  
  - ISO 19131 on specifications,
  
  - ISO 19115/19139 – metadata
  
- Currently being revised and combined.



# Benefits of ISO 19100

- Common understanding and terminology
- Facilitates interoperability
- Reduces duplication of software development and future-proofs investment
- Compatibility with other standards (e.g. metadata) is managed ISO at the standards level
- Provides a well recognised and trusted source of reference
- Robust change management process

# Why implement ISO 19100?

- Demonstrating data quality increasingly critical
- Ensuring fitness for purpose for existing users
- A “good fit” for the NLPG and current objectives
- Based on established best practice
- NLPG increasingly being discovered by new, international audiences
- Legislative/Industry drivers - INSPIRE & UKLP



# The ISO Data Quality Evaluation Process

Figure 2.2 Data Quality Evaluation Process



Guidelines for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies

Antti Jakobsson, Jørgen Giversen  
[www.eurogeographics.org](http://www.eurogeographics.org)



# Data Quality Elements

- ISO 19113 defines the quality of a geographic data product in terms of the following *data quality elements*:
  - Completeness
  - Logical Consistency
  - Topological Consistency
  - Positional Accuracy
  - Temporal Accuracy
  - Thematic Accuracy



# Example NLPG Data Quality Element

- **Element**
  - Completeness
- **Subelement**
  - Omission
- **Scope**
  - “ Current residential property in England and Wales”



# The value of measurement

- “When you measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind“

– Lord Kelvin



# Measuring Data Quality

- The best measures are:
  - Justified – i.e. tied to strategy, specification or requirement
  - Well defined
  - Understood and agreed by stakeholders
  - Not subjective
  - Measurable!
  - Quantitative



# Writing Measures

- "The idea is for each requirement to have a quality measure that makes it possible to divide all solutions to the requirement into two classes: those for which we agree that they fit the requirement and those for which we agree they do not fit the requirement"
  - Christopher Alexander, Notes on the Synthesis of Form



# Example NLPG Data Quality Measure

- **Element**
  - Completeness
- **Subelement**
  - Omission
- **Scope**
  - “Current residential property in England and Wales”
- **Measure**
  - “99% known residential objects exist in the dataset”



# Example NLPG Data Quality Evaluation Method

- **Element**
  - Completeness
- **Subelement**
  - Omission
- **Scope**
  - “Current residential property in England and Wales”
- **Measure**
  - “99% known residential objects exist in the dataset”
- **Evaluation Method**
  - **Comparison with VOA Council Tax data**



# Example Data Quality Result

- **Result**
  - Accepted
- **Value type**
  - Boolean
- **Date**
- 17 May 2010



# Monitoring and Reporting

- Data Quality Information can be reported as:
  - Metadata (ISO 19119 & 19139, INSPIRE, GEMINI 2)
  - Data Quality Report (Annex II of ISO 19114)



# Importance of context

- Statistics are meaningless, ambiguous and or misunderstood if not reported with context
- Using the ISO standard encourages better use of context
- 3 genuine interpretations of the above result:
  - “The national processes are not working as well as necessary”
  - “The local processes are working very well”
  - “The data product meets a defined user requirement”

# What do I do with it?

- Publish to aid discovery of data
- Provide to users to demonstrate fitness for purpose and provide confidence
- Include in tender response
- Improve process

