

# Metadata & the NPHL

Peter Cornelissen  
Head of IM&T, ERPHO

# What is metadata?

- Most commonly described as ‘Data about data’
- It is the ‘who, what, where and when’ of a resource
- As far as ‘we’ are concerned, it is more specifically ‘structured information about a resource’

# Why do we need it?

- It puts labels on a resource to tell us its title, subject, type, creator, date of publication, etc.
- It describes, explains, locates, or otherwise makes it easier to manage a resource
- It improves the chances of the resource being found by the right person at the right time
- Maintains consistency across public sector organisations (both locally and internationally)

# Why do we need it?



It helps interoperability by providing a common set of descriptors even if underneath you use different systems (XML)

# Where does it come from?

- Dublin Core (1995)
- eGMS (electronic Government Metadata Standard)
- APHO
  - have adopted e-GMS, for all their web sites
  - for interoperability, we have defined a subset of e-GMS as well as encoding scheme/s

# How can we use metadata?

- We can describe **ALL** our resources with it
  - When we add any resources to our web site
  - We use the same metadata elements for each resource
- We can use it when we search for resources
  - We can search for any of the metadata elements
  - We can use the NPHL for the subject

# Why not a search engine?

- Search engines have no structure
- They only look for keywords (currently)
  - If the title, description or meta tags do not contain the words you are searching with, you won't find it
- Its results are unstructured
  - You cannot get serendipitous results
  - You can get too many results

- And you can end up fishing around aimlessly



# The e-GMS elements

- Accessibility - Mandatory
- Audience - Optional
- Contributor - Optional
- Coverage - Recommended
- Creator - Mandatory
- Date - Mandatory
- Description - Recommended
- Disposal - Optional
- Format – Optional
- Identifier – Mandatory
- Language - Recommended
- Publisher - Mandatory
- Relation - Optional
- Rights - Optional
- Source - Optional
- Status - Optional
- Subject – Mandatory
- Title – Mandatory
- Type – Mandatory

# Example - Refinements for the element 'Date'

- Acquired
- Available
- Created
- Cut-off date
- Closed
- Date accepted
- Date copyrighted
- Date submitted
- Declared
- Issued
- Modified
- Next version due
- Updating frequency
- Valid

- Q: Can we add our own elements or refinements?
- A: Yes
  - They may only be useful on your own systems
  - They can add more useful information (so why not try to make them standard?)

# The 'Type' element (and 'PHRTES')

- Dublin Core recommends the use of a controlled vocabulary for the Type element
  - APHO have developed its own called the Public Health Resource Type Encoding Scheme (PHRTES)
    - This has top level types, most of which have their own refinements
    - To be managed by the same working group as the NPHL
- 
- |              |                |
|--------------|----------------|
| – Dataset    | – Presentation |
| – Collection | – Method       |
| – Table      | – Event        |
| – Map        | – News         |
| – Chart      | – Contact      |
| – Report     | – Website      |
| – Discussion | – Media        |

# The 'Subject' element

- In Dublin Core 'The topic of the resource'
- In eGMS, it is described as 'A topic of the content of the resource'
- It can be
  - Keywords
  - Categories
  - key phrases
  - Classification codes
- Best practice is to use a controlled vocabulary

# 'Subject' element controlled vocabularies

- Integrated Public Sector Vocabulary
  - A high level taxonomy that covers ALL (UK) Government and public areas
- NPHL – National Public Health Language
  - A poly-hierarchical taxonomy that covers terms used by Public Health professionals
  - Approx. 3500 terms
- SNOMED CT
  - A poly-hierarchical taxonomy for Clinical Terms
  - More than ½ million terms
- Many others (MESH, ICD10 etc.)

# National Public Health Language

- 2002 - PHIT (Public Health Information Thesaurus) - HDA/NICE
- 2002 - PHITS (Public Health Information Tagging Standard) – ERPHO/APHO
- 2004 - PHIT + PHITS = NPHL (National Public Health Language)
- 2007 - SNOMED CT + NPHL = PHT? (Public Health Taxonomy?)

# What's been happening?



Things have been quite slow

# What's been happening?

- **Steering Group (2003)**
  - Assembled from volunteers from:
    - APHO
    - NICE
    - DH
    - HPA
    - PCTs
  - Manages the policies and procedures
  - Has final say on releases
  - Attempts to get funding
- **Editorial board (2004)**
  - Smaller group (currently members from APHO, DH & HPA)
  - Responsible for updating language
- There has been 3 releases to date (now at V1.2)

# What is happening now?

- We have finally got funding from DH through the I&I Strategy and project is proceeding
  - It is part of the National Library for Public Health (NLPH, shame about the acronym!!!)
- The NPHL has been updated to version 1.2 and has been submitted this to SNOMED for first pass mapping
  - A first pass has been done and about 50% of terms match
- The Steering Group and Editorial Board are morphing into a Working Group
- The Working Group take on other Metadata (esp. PHRTES)

# What is going to happen?

- It will change its name (Public Health Thesaurus – PHT?)
- It will become the PH sub-set of SNOMED CT
- It will become the de facto taxonomy for the NLPH as well as all other in Public Health
- It will have a permanent management structure
- The Working Group will include members from APHO, DH, HPA, NICE, PCTs, CfH, SNOMED CT, BMJPH and will (?) be responsible for recommended PH metadata standards

# What must happen?



We must work together to improve

# What 'must' happen?

- Set up smaller working groups to develop areas of the NPHL
- Develop training resources and a training programme to enable NPHL to be used to search for and tag information resources and commence a training programme.
- Add 100s of synonyms to make the NPHL a true thesaurus
- Create 'Views' of the NPHL for specific purposes (e.g. as with the NLPH)

# What 'should' happen?

- Develop training resources and a training programme
  - to enable NPHL to be used to search for and tag information resources and commence a training programme.
- Establish and service an active web-based User Forum
  - which would propose changes and new terms to the Editorial Board and ensure that NPHL is meeting users' needs.
- Continue and strengthen links with the Public Health Classifications project in Australia.
- Establish permanent links with:
  - EU Public Health Portal.
  - MeSH and its associated Unified Medical Language System

# The NPHL to SNOMED mapping project

# Current NPHL Top Terms

1. Health, Public Health, Health Promotion
2. Determinants of Health
3. Death, Disease and Disability
4. People and Populations
5. Settings and Places
6. Health Services and their Management
7. Public Health Methods, Theory and Research
8. Communication and Knowledge
9. Time Factors
10. Equipment

# NLPH 'View' Top Terms

- People and Populations
- Places
- Risks and Determinants [upstream]
- Outcomes and Diseases [downstream]
- Methods, Interventions and Services
- Infrastructure

# SNOMED CT - High level categories

1. **clinical findings**
2. **procedures**
3. observable entities
4. body structures (anatomical, morphological)
5. organisms
6. pharmaceutical/biologic products (drugs, etc)
7. substances ("matter", not individuals)
8. physical objects (devices, etc)
9. physical forces
10. events
11. specimens
12. environments & geographic locations
13. social contexts
14. context-dependent categories
15. qualifier values
16. staging and scales
17. special concepts
18. linkage concepts
19. SNOMED CT UK administrative concepts

# SNOMED CT - Fully Specified Name tags

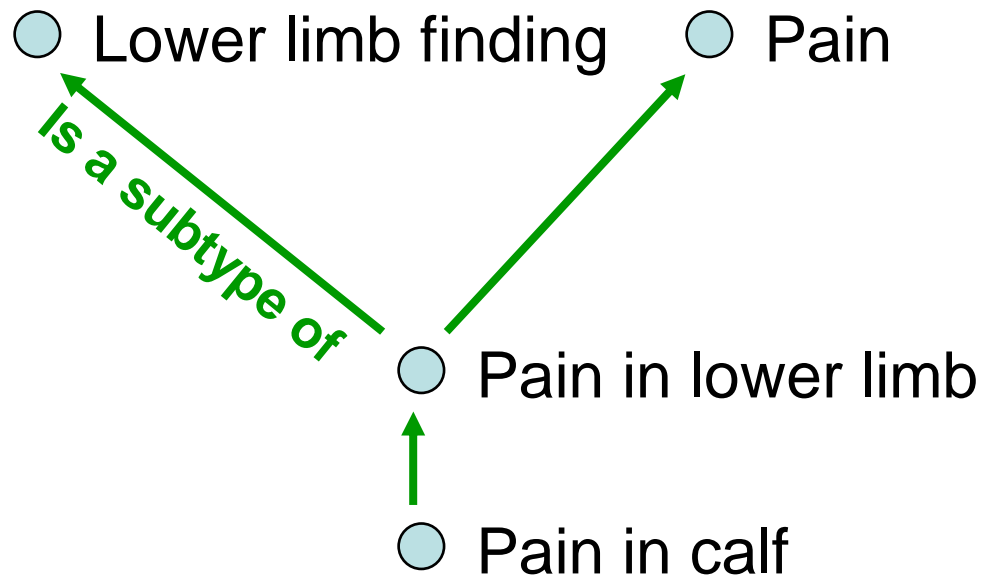
emergency appendicectomy (procedure) 174036004

- Finding
- Disorder
- Procedure
- Organism
- Substance
- Product
- Physical force
- Physical object
- Attribute
- Person
- Social concept
- Religion/philosophy
- Lifestyle
- Body structure
- Morphological abnormality
- Context-dependent category
- Environment
- Geographic location
- Qualifier value
- Event
- Observable entity (many types here)
- Navigational concept
- Namespace concept
- Inactive concept
- Tumor staging

# SNOMED CT - Subtype relationships

- Every concept is a refined type of one or more other concepts
- For example
  - “Pain in the leg” is a type of “pain”
  - “Pain in the leg” is a type of “lower limb finding”
- SNOMED CT represents these defining relationships with the relationship type “is a”

# SNOMED CT - Subtype relationships



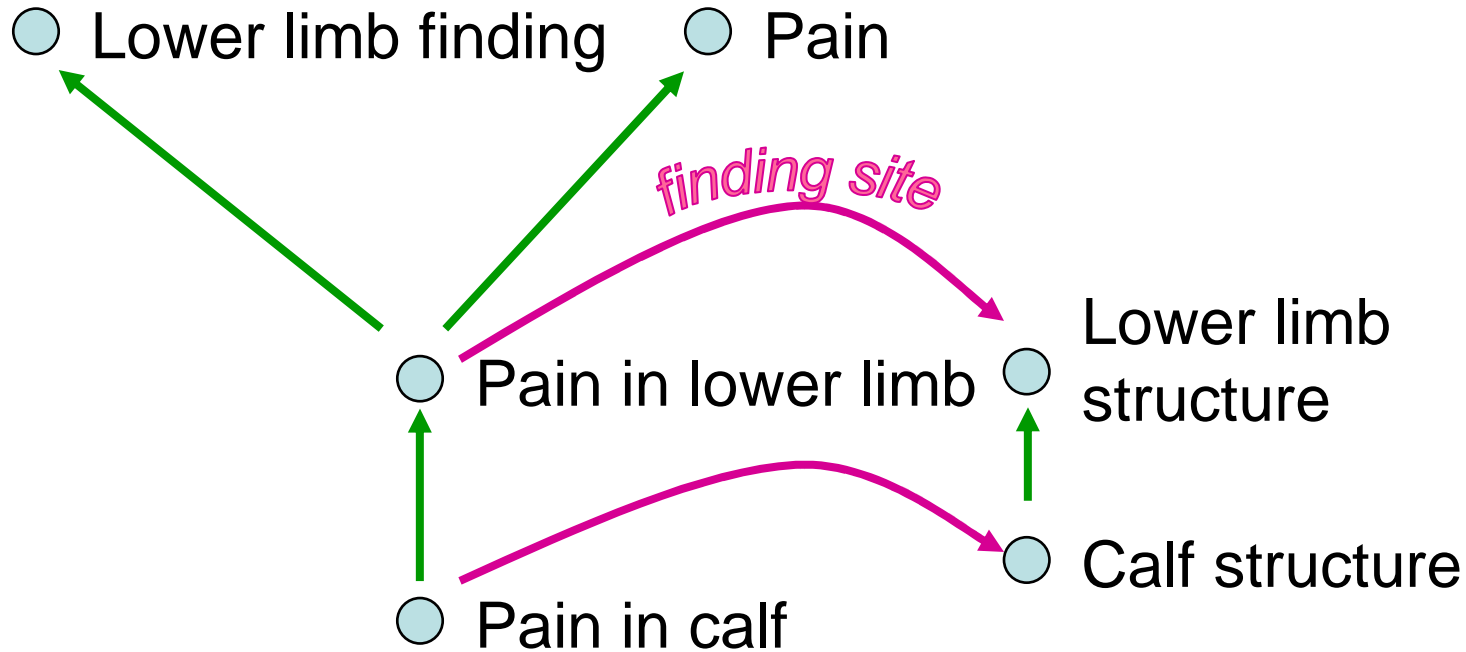
A pain in the calf *is-a* pain in the lower limb, and  
Pain in the lower limb *is-a* pain, and *is-a* lower limb finding



# Other defining relationships

- The difference between two concepts may be represented by other defining relationships
  - Only relationships that are necessarily true are defining relationships
- For example
  - “Pain in calf”
    - ... is distinguished from “pain in lower limb” by the defining relationship ...
  - “finding site” = “calf structure”

# Other defining relationships



A pain in the calf *has finding site* calf

Pain in the lower limb *has finding site* lower limb

# Primitive & fully-defined concepts

- A concept is “fully defined”
  - if its definition is sufficient to distinguish it from all its supertype concepts
  - may acquire new (inferred) super-types and sub-types as a consequence of the classification process
- A concept is “primitive”
  - if it is not “fully defined”
  - may acquire new (inferred) super-types as a consequence of the classification process

# SNOMED CT - Primitive & fully defined concepts

- Head injury
  - Is a = Disease
  - Associated morphology = Traumatic abnormality
  - Finding site = Head structure
  - Fully Defined
- Aching pain
  - Is a = Pain
  - Primitive
- Headache
  - Is a = Aching pain
  - Finding site = Head structure
  - Fully Defined

The end