

# Using the Mauro Data Mapper to represent phenotypes

**James Welch, Jim Davies**

*University of Oxford*



# Introduction to Mauro

A “Metadata Catalogue” for modeling data assets, and data specifications

Focused on data semantics, and interoperability

Also interested in capturing data flows, summary metadata, and usage

Open-source tool, used in HDR UK Innovation Gateway

The screenshot shows the Mauro Data Mapper (MDM) website. The header is dark blue with the MDM logo (MAURO DATA MAPPER) on the left, navigation links (Home, Browse models, Search, About) in the center, and a user profile (James Welch, Administrator) on the right. The main content area has a dark blue background with white text. It features a headline: "Use the Mauro Data Mapper platform to create shared documentation for your data, and to collaborate on the definition of new data models". Below this is a sub-headline: "Automatically import your existing schemas; link, annotate and share them; use these definitions in the creation of new software components." and a paragraph: "Mauro was previously known as the Metadata Catalogue, and has been built at the University of Oxford with support from the National Institute for Health Research, and NHS Digital." A dark blue button labeled "Browse our Models" is positioned below the text. To the right of the text is an illustration of a computer monitor displaying a share icon, with a dashed line connecting it to a vertical stack of three green bar charts, each with a blue circle to its right. Below the main content area is a light gray section titled "Benefits of using the Mauro Data Mapper" with the word "Features" above it. This section contains three columns, each with an icon, a title, and a description. The first column has a green circular icon with a white recycling symbol, the title "Automate", and the description: "Automatically create data models from existing artefacts, such as relational databases, Excel spreadsheets, XML Schema, or definitions in OWL or UML. Visualise your data to check for completeness and consistency." The second column has a blue circular icon with a white balance scale, the title "Collaborate", and the description: "Create new models, re-using existing definitions. Manage versioning and publication lifecycles, and comment directly on parts of a model." The third column has an orange circular icon with a white gear and a plus sign, the title "Generate", and the description: "Create new software and configuration from existing definitions - generate forms, websites and databases - and make your documentation go further using our extensive APIs."

**MDM** MAURO DATA MAPPER

[Home](#) [Browse models](#) [Search](#) [About](#)

James Welch Administrator

Use the Mauro Data Mapper platform to create shared documentation for your data, and to collaborate on the definition of new data models

Automatically import your existing schemas; link, annotate and share them; use these definitions in the creation of new software components.

Mauro was previously known as the Metadata Catalogue, and has been built at the University of Oxford with support from the National Institute for Health Research, and NHS Digital.

[Browse our Models](#)

Features

**Benefits of using the Mauro Data Mapper**

**Automate**

Automatically create data models from existing artefacts, such as relational databases, Excel spreadsheets, XML Schema, or definitions in OWL or UML. Visualise your data to check for completeness and consistency.

**Collaborate**

Create new models, re-using existing definitions. Manage versioning and publication lifecycles, and comment directly on parts of a model.

**Generate**

Create new software and configuration from existing definitions - generate forms, websites and databases - and make your documentation go further using our extensive APIs.

# Data models

The screenshot displays a web-based data model editor. On the left is a sidebar with a tree view of models, including 'Core information standard', 'About me', 'Additional support plans', 'Admission details', 'Admission details record entry' (selected), 'Individual accompanying person', 'Relationship', 'Alerts', 'Allergies and adverse reactions', 'Assessments', 'Care and support plan', 'Contacts with professionals', 'Contingency plans', 'Current pregnancy', 'Discharge details', 'Documents (including correspondence and images)', 'End of life care', 'Equipment and adaptations', 'Examination findings', 'Family history', 'Formulation', 'Future appointments', 'GP practice', 'Individual requirements', 'Investigation results', 'Investigations requested', 'Legal information', 'Medications and medical devices', 'Participation in research', 'Person demographics', 'Personal contacts', 'Plan and requested actions', 'Primary support reason', 'Problem list', 'Procedures and therapies', and 'Professional contacts'.

The main panel shows the 'Admission details record entry' data class. It includes a search bar, a list of models, and a table of properties. The table has columns for 'Name', 'Details', and 'Multiplicity'. The data classes listed are:

Name	Details	Multiplicity
Admission method <a href="#">DataClass</a>	Description How the person was admitted to hospital e.g. elective, emergency, maternity, transfer etc.	0..1
Admitted to <a href="#">DataClass</a>	Description The hospital the person was admitted to.	0..1
Individual accompanying person <a href="#">DataClass</a>	Description Details of the accompanying individual and the extent to which they have provided the information... <a href="#">Show more</a>	0..1
Reason for admission <a href="#">DataClass</a>	Description The health problems and issues experienced by the person that prompted the decision to admit to... <a href="#">Show more</a>	0..1
Responsible consultant <a href="#">DataClass</a>	Description The consultant who has overall responsibility for the person (may not actually see the person) <a href="#">Show more</a>	0..1
Source of admission <a href="#">DataClass</a>	Description Where the person was immediately prior to admission, e.g. usual place of residence, temporary place... <a href="#">Show more</a>	0..1
Specialty <a href="#">DataClass</a>	Description The specialty e.g. physiotherapy, oncology, mental health etc	0..1

High-level information about a data model itself

Schema details – structured descriptions of classes and elements

Multiplicities, data type constraints, references to other models.

Re-use of definitions to encourage standardisation

# Terminologies

A terminology is a first-class model within Mauro

Unstructured collections of terms, and relationships between them.

Typically hierarchical, but supporting arbitrary relationship types too

Browse, search functionality built into the interface

Can be used as target for semantic links, or as a data type within a model

-  International Classification of Diseases (ICD) Version 10 Edition 5 |  1.0.0
- I: Certain infectious and parasitic diseases
  - + A00-A09: Intestinal infectious diseases
  - A15-A19: Tuberculosis
    - A15: Respiratory tuberculosis, bacteriologically and histologically confirmed
      - A15.0: Tuberculosis of lung, confirmed by sputum microscopy with or without culture
      - A15.1: Tuberculosis of lung, confirmed by culture only
      - A15.2: Tuberculosis of lung, confirmed histologically
      - A15.3: Tuberculosis of lung, confirmed by unspecified means
      - A15.4: Tuberculosis of intrathoracic lymph nodes, confirmed bacteriologically and histologically
      - A15.5: Tuberculosis of larynx, trachea and bronchus, confirmed bacteriologically and histologically
      - A15.6: Tuberculous pleurisy, confirmed bacteriologically and histologically
      - A15.7: Primary respiratory tuberculosis, confirmed bacteriologically and histologically
      - A15.8: Other respiratory tuberculosis, confirmed bacteriologically and histologically
      - A15.9: Respiratory tuberculosis unspecified, confirmed bacteriologically and histologically
    - A16: Respiratory tuberculosis, not confirmed bacteriologically or histologically
      - A16.0: Tuberculosis of lung, bacteriologically and histologically negative
      - A16.1: Tuberculosis of lung, bacteriological and histological examination not done
      - A16.2: Tuberculosis of lung, without mention of bacteriological or histological confirmation
      - A16.3: Tuberculosis of intrathoracic lymph nodes, without mention of bacteriological or histological confirmation
      - A16.4: Tuberculosis of larynx, trachea and bronchus, without mention of bacteriological or histological confirmation
      - A16.5: Tuberculous pleurisy, without mention of bacteriological or histological confirmation
      - A16.7: Primary respiratory tuberculosis without mention of bacteriological or histological confirmation
      - A16.8: Other respiratory tuberculosis, without mention of bacteriological or histological confirmation
      - A16.9: Respiratory tuberculosis unspecified, without mention of bacteriological or histological confirmation
  - + A17: Tuberculosis of nervous system
  - + A18: Tuberculosis of other organs
  - + A19: Miliary tuberculosis
- + A20-A28: Certain zoonotic bacterial diseases
- + A30-A49: Other bacterial diseases
- + A50-A64: Infections with a predominantly sexual mode of transmission
- + A65-A69: Other spirochaetal diseases
- + A70-A74: Other diseases caused by chlamydiae
- + A75-A79: Rickettsioses
- + A80-A89: Viral infections of the central nervous system
- + A92-A99: Arthropod-borne viral fevers and viral haemorrhagic fevers

# Code Sets

Pneumonia Secondary Care

Draft

Item type:

Code Set

Branch:

main

Documentation Version:

1.0.0

Description

Terms (15)

Comments

History

Links

Attachments

Rules (0)

Terms

15

+ Add Term

Terminology	Term	Definition
International Classification of Diseases (ICD) Version 10 Edition 5	B01.2: Varicella pneumoniaJ17.1	Varicella pneumoniaJ17.1
International Classification of Diseases (ICD) Version 10 Edition 5	B05.2: Measles complicated by pneumoniaJ17.1	Measles complicated by pneumoniaJ17.1
International Classification of Diseases (ICD) Version 10 Edition 5	B20.6: HIV disease resulting in Pneumocystis jirovecii pneumonia	HIV disease resulting in Pneumocystis jirovecii pneumonia
International Classification of Diseases (ICD) Version 10 Edition 5	B25.0: Cytomegaloviral pneumonitisJ17.1	Cytomegaloviral pneumonitisJ17.1
International Classification of Diseases (ICD) Version 10 Edition 5	J10.0: Influenza with pneumonia, seasonal influenza virus identified	Influenza with pneumonia, seasonal influenza virus identified
International Classification of Diseases (ICD) Version 10 Edition 5	J11.0: Influenza with pneumonia, virus not identified	Influenza with pneumonia, virus not identified
International Classification of Diseases (ICD) Version 10 Edition 5	J12: Viral pneumonia, not elsewhere classified	Viral pneumonia, not elsewhere classified
International Classification of Diseases (ICD) Version 10 Edition 5	J13: Pneumonia due to Streptococcus pneumoniae	Pneumonia due to Streptococcus pneumoniae
International Classification of Diseases (ICD) Version 10 Edition 5	J14: Pneumonia due to Haemophilus influenzae	Pneumonia due to Haemophilus influenzae
International Classification of Diseases (ICD) Version 10 Edition 5	J15: Bacterial pneumonia, not elsewhere classified	Bacterial pneumonia, not elsewhere classified
International Classification of Diseases (ICD) Version 10 Edition 5	J16: Pneumonia due to other infectious organisms, not elsewhere classified	Pneumonia due to other infectious organisms, not elsewhere classified
International Classification of Diseases (ICD) Version 10 Edition 5	J17: Pneumonia in diseases classified elsewhere	Pneumonia in diseases classified elsewhere

A new model type, collating a set of terms from one or more terminologies

Can be used to define phenotypes or cohorts within datasets

Usage information to define applicability of terms within terminologies

Can be used as target for semantic links, or as a data type within a model

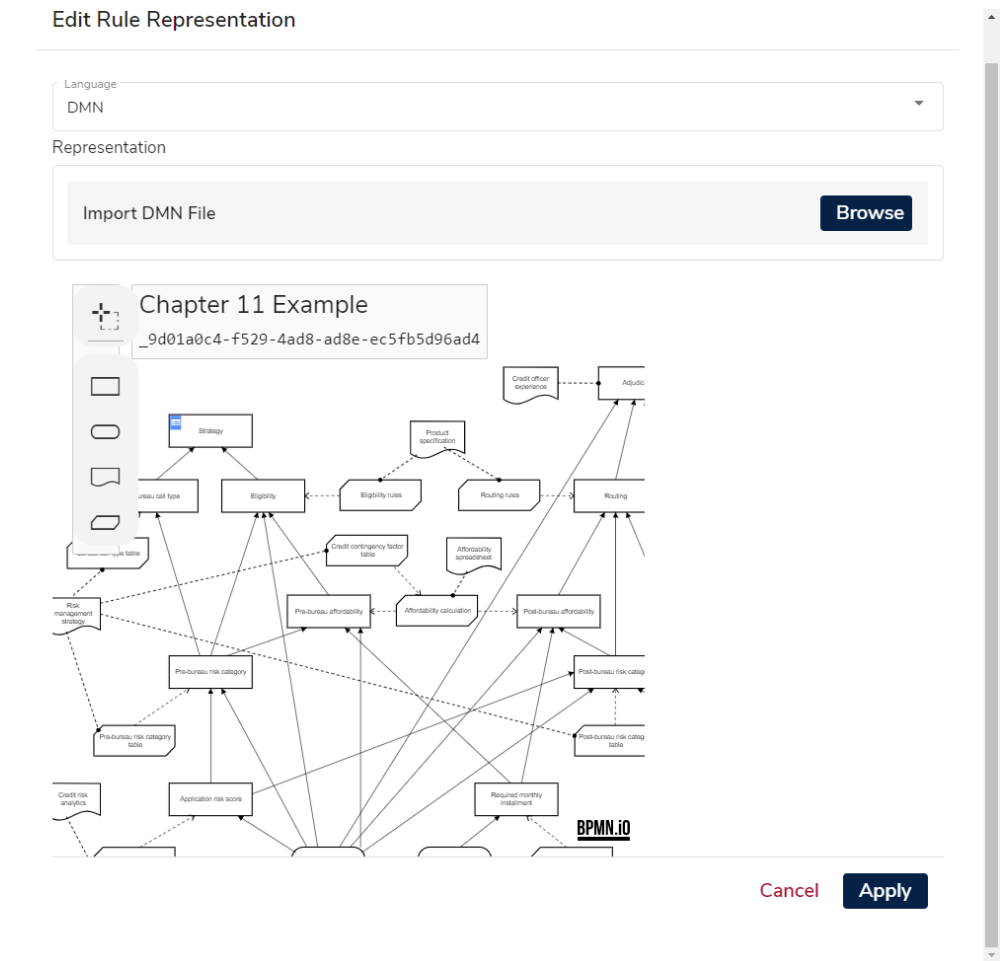
# Rules

Any model, or model component, can define rules, or constraints

Each named rule may have a number of representations in different languages

Current support for SQL, DMN, Java, Python.

These can be used for documenting constraints, validation, or profiling data.



# Iterative, collaborative development

Support for versioning – lightweight form of GitFlow allowing forking and merging

Tools for collaborative editing

Publication life-cycle, including finalization and superseding

Variety of import / export formats

Federation through publish / subscribe mechanism

# FHIR integration

Ongoing work with NHS Digital -  
supporting NHS Data Dictionary

Integration with fhir.nhs.uk and  
new NHS terminology server

‘CodeSystem’ -> Terminology

‘ValueSet’ -> CodeSet

‘StructureDefinition’ -> DataModel

All linked to the FHIR Core Model,  
imported via XSD

FHIR 3.0.1: DiagnosticReport		
FHIR 3.0.1		
DescriptionElementsContextHistoryDataRules		
Data Classes 3		
Name	Details	Multiplicity
diagnosticReportStatus DataClass	Description: The status of the diagnostic report as a whole.	
diagnosticReport.Performer DataClass	Description: The findings and interpretation of diagnostic tests performed on patients, groups of patients,... <a href="#">Show more</a>	
diagnosticReport.Image DataClass	Description: The findings and interpretation of diagnostic tests performed on patients, groups of patients,... <a href="#">Show more</a>	
Data Elements 19		
Name	Details	Multiplicity
image DataElement	Description: A list of key images associated with this report. The images are generally created during the... <a href="#">Show more</a>  Data Type: <a href="#">diagnosticReport.Image</a> (Reference) [ref: diagnosticReport.Image]	0..*
text DataElement	Description: A human-readable narrative that contains a summary of the resource, and may be used to represent... <a href="#">Show more</a>  Data Type: <a href="#">narrative</a> (Reference) [ref: narrative]	0..1



# Access

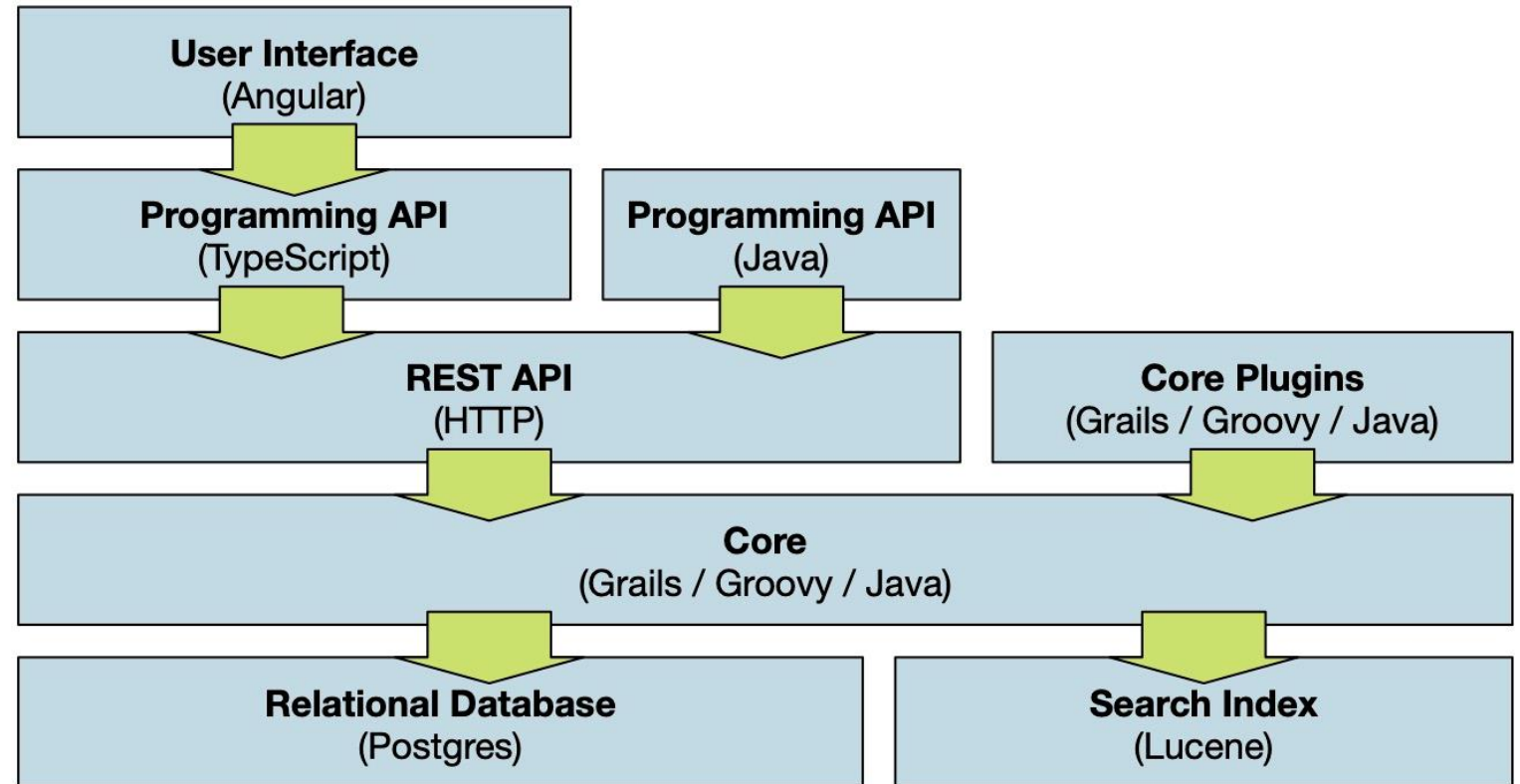
REST API conforming to standard principles

Client libraries defined in Java / Groovy, Javascript / Typescript, and (new) .NET. Python on the way...

Web interface for browsing / discovery

Plugins built to import / export from other data formats

Looking to increase support for federation with non-Mauro sources



# Standardisation / integration

Interchange formats, but also APIs

Import of Caliber phenotypes from GitHub

Import / export to NHS Digital Terminology Server

SPARQL endpoint for semantic web