UNIVERSITY of York

Domain-specific modeling languages for Citizen Science data provenance

MDENet seedcorn fund project

MDENet | The home of model-driven engineering

Aston University
BIRMINGHAM UK

Research Metadata

- FAIR Principles
  - **F**indability, **A**ccessibility, **I**nteroperability and **R**euse
- Reusable datasets
  - Documenting the creation encouragers use/reuse of datasets
  - User of a dataset is given an explanation of the creation process
- Citizen Science
  - Can be at a disadvantage lacking metadata expertise
  - Data produced is often mistrusted

  Example Citizen Science project: https://naturehood.uk/survey-your-space

Approaches to lineage documentation

NONE - Do nothing!

- “My data explains itself!”
- Lack of metadata knowledge or training

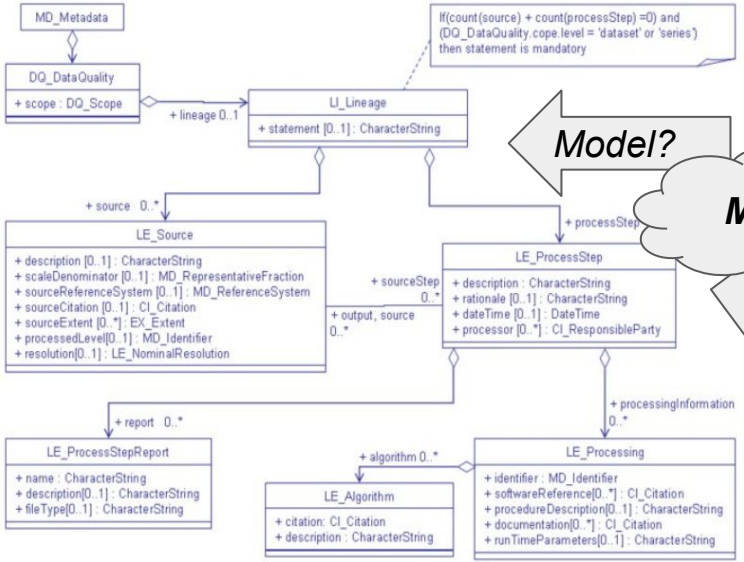
Unstructured

- Just write a document
- Not machine parsable
- May not follow a “standard”

Structured

- Lineage specifications
  - ISO19115-3
- Minimum level of detail
- Interchange formats (XML)
- Complex and difficult
  - Non-experts will struggle

ISO9115-3: UML, XML, Tools

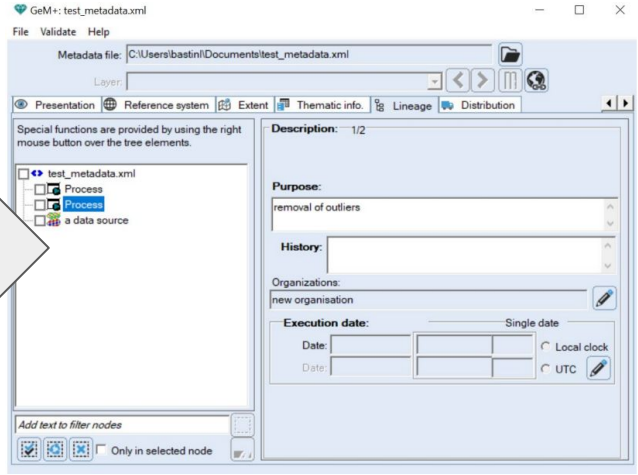


Model?

MDE?

Model editor?

Model-to-Text?




```
<gmd:dataQualityInfo>
  <gmd:DQ_DataQuality>
    <gmd:scope>
      <gmd:DQ_Scope>
        <gmd:level>
          <gmd:MD_ScopeCode
          codeList="http://www.isotc211.org/2005/resources/codeList.xml#MD_ScopeCode"
          codeListValue="dataset"/>
        </gmd:level>
      </gmd:DQ_Scope>
    </gmd:scope>
    <gmd:lineage>
      <gmd:LI_Lineage>
        <gmd:statement>
          <gco:CharacterString>Due to the map generation method, the quality of the map can never be
          uniform. The overall quality of the map depends heavily on the individual quality of the
          data for the different countries.</gco:CharacterString>
        </gmd:statement>
      </gmd:LI_Lineage>
    </gmd:lineage>
  </gmd:DQ_DataQuality>
</gmd:dataQualityInfo>
```

Model-driven solution

Aim: try and reduce the barrier of entry to creating lineage documentation that meets specifications like ISO19115-3.

- Modeling language (based on ISO19115-3)
  - Use generators for model editors and text file outputs
  - One example of file output: the standard XML files
- Assistive tools with a “text-like” editor (MPS, Abstract Syntax Tree)
  - Form-like structure, task automation, auto-complete, hints
- Model your project once, produce multiple deliverables

Project: MPS Prototype

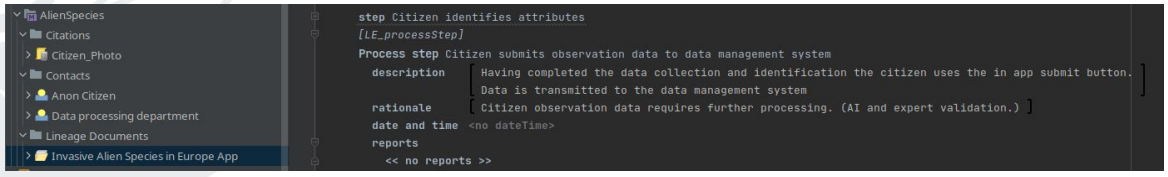
Project participants:

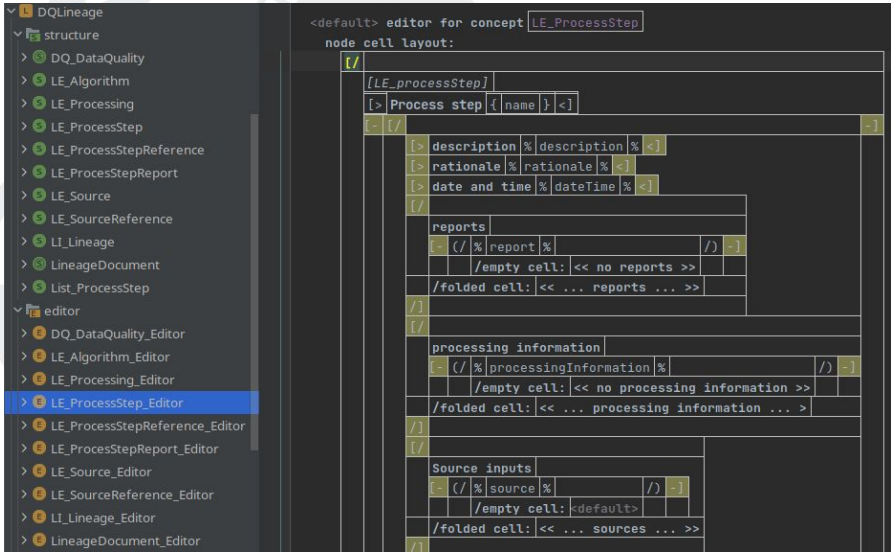
- Aston University
  - Owen Reynolds, *PhD student*
  - Lucy Bastin,
    Research metadata & Citizen Science expert
- University of York
  - Antonio Garcia-Dominguez, *MDE expert*
- Earthwatch
  - James Sprinks, *Citizen Science expert*

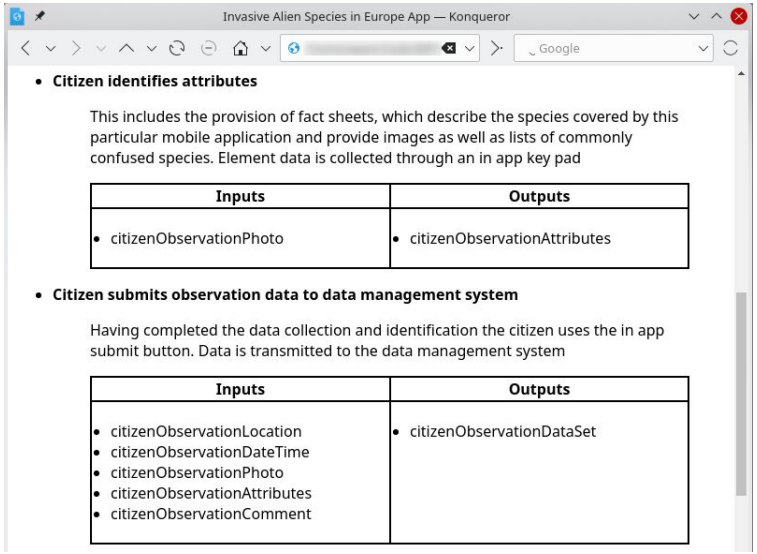
Why Jetbrains MPS?

- Open-source language workbench
  - Structure, Editor, Type system
  - Transformations, Constraints
- Projectional editor
  - Quickly implements complex
    “text-like editor with a guide
  - Looks like a document but has cells
  - Intentions & auto-complete
  - Automated validation of user inputs

MPS: ISO 19115-3







Next step: User testing

Earthwatch Citizen Science project managers

- What do Citizen Science project managers think of the modelling style approach to project documentation?
- Is the “text-like” editor interface helpful/intuitive?
- What other kinds of output formats would helpful?

Observations so far…

- Using ISO 19115-3 for documentation is hard
  - Overwhelming with the details
  - Unclear documentation, which is also overwhelmingly big!
- Model representations can be tuned to audiences
  - One language model and one project model
  - Multiple editors and outputs
- Modelling tools and lineage specifications share a problem
  - Both become complicated with the levels of detail involved
  - Higher levels of abstraction and assistive tools help

Future work

- User feedback will guide
  - Additional output types (graphical representations of process flows?)
  - Editor experience (different projectional editors)
- Tool delivery via the web
  - “Cloud based” project modelling tool
  - Remove the requirements for installing software
  - Enable collaborative modelling for project documentation
- Follow up research grant application
  - To build a set of standards-based notations for the Citizen Science community using MDE tools to support the development process

Thank you for attending!

Any questions?

Contact details:

Antonio Garcia-Dominguez
a.garcia-dominguez@york.ac.uk

Lucy Bastin
l.bastin@aston.ac.uk

Owen Reynolds
180200041@aston.ac.uk