CI Pipeline for generating Linked Open Data ontology documentation

Speaker: Kirill Bulert, Lars-Peter Meyer
InfAI e.V.
Introducing Project StahlDigital

- BMBF funded 2021-2024.
- Partners: MPIE, Fraunhofer IWM, InfAI
- Part of Plattform MaterialDigital
  - Different materials, e.g. steel, copper, rubber, glass, concrete ...
  - Different problems, e.g. simulation workflows, digital twin, ML, ...
  - All using ontologies and tackle material science challenges
- Main Topics Project StahlDigital
  - ontology development workflow
  - development of domain specific ontology
  - Ontology based workflows

See also: https://material-digital.de/project/6
Challenges

- Building domain specific ontologies is hard
  - Domain experts lack knowledge about ontologies and tooling
  - Ontology experts lack knowledge about domain
- Publishing ontologies is also hard
  - Manual approach is error-prone
  - Ontologies require updates
    - Different representations required for humans and machines
- How to tackle this challenges cost efficient, especially if projects run out of funding?
  - Domains and servers need to be kept alive
  - Many 404s around
  - Dublin Core, FOAF and SKOS are around for over 20 years

See also:
https://purl.org
https://w3id.org
https://archivo.dbpedia.org
Approach chosen by StahlDigital: Spreadsheet as input

- Input interface for terms and data
- Ontology and Domain experts both used to work with
- Software available on most computers
- Data generally stored in spreadsheets within our domain

Special thanks to Eccenca GmbH for initial idea and input
https://eccenca.com
Approach chosen by StahlDigital: Git repository and CI pipeline

- Versioned storage
- Pipeline with input conversion, tests, publishing
- Tools used: OnlyOffice, Gitlab CI, Nextflow, YARRRML, JekyllRDF, ...
Approach chosen by StahlDigital: publishing with Github & w3id.org

- Documentation stored on Github
- Robust content aware redirection via w3id.org
W3id Redirect

- W3id utilizes Apache's httpd content-type aware rewrite rules
- Redirect (Code 303, see Other) based on URL and content type as suggested by “Cool URIs for the Semantic Web”

See also:
https://www.w3.org/TR/cooluris/
W3id Redirect

- W3id utilizes Apache's httpd content-type aware rewrite rules
- Redirect(Code 303) based on URL and content type as suggested by “Cool URIs for the Semantic Web”

See also: https://www.w3.org/TR/cooluris/
### Further work and summary

<table>
<thead>
<tr>
<th>TODO</th>
<th>Doing</th>
<th>Done</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxy instead of redirect</td>
<td>Better tooling for testing w3id redirects</td>
<td>Ontology input via spreadsheet</td>
</tr>
<tr>
<td>Pipeline modules: Ontoflow</td>
<td>Add more Pipeline modules</td>
<td>Automated development pipeline</td>
</tr>
<tr>
<td>improve w3id update workflow</td>
<td>Ontology development</td>
<td>Low maintenance infrastructure</td>
</tr>
<tr>
<td>Evaluation of our approach</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

See also: [https://gitlab.com/infai/ontoflow](https://gitlab.com/infai/ontoflow)
Thank you for your attention

Questions?

Contact:
- Kirill Bulert: bulert@infai.org
- Lars-Peter Meyer: lpmeyer@infai.org