

Building Ontologies for Reuse

Lessons Learned from Unit Ontologies

Sirko Schindler¹ and Jan Martin Keil²

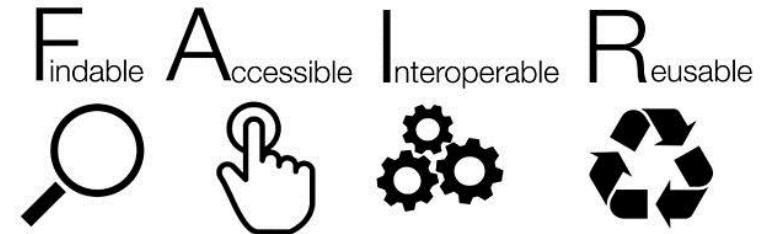
¹*DLR Institute of Data Science, Jena, Germany*

²*Friedrich Schiller University Jena, Germany*



Motivation

- Reuse - a core ontology promise
 - Linked Data Cloud
- Different issues hindering reuse of existing ontologies
 - Redundant development
- Units are a common part of many ontologies
 - Multiple ontologies available
 - Reuse has its pitfalls ...



Semantic Web 10 (2019) 03–51
DOI 10.3233/SW-180310
IOS Press

33

Comparison and evaluation of ontologies for units of measurement

Jan Martin Keil^{a,*} and Sirkо Schindler^b
^aHeinz Nixdorf Chair for Distributed Information Systems, Institute for Computer Science, Friedrich Schiller University Jena, Germany
E-mail: jan-martin.keil@uni-jena.de; ORCID: <https://orcid.org/0000-0002-7733-0193>

^bHeinz Nixdorf Chair for Distributed Information Systems, Institute for Computer Science, Friedrich Schiller University Jena, Germany
E-mail: sirkо.schindler@uni-jena.de; ORCID: <https://orcid.org/0000-0002-0964-4457>

Editor: Boyce Bruton, Geological Survey of Canada, Canada
Solicited reviews: Steve Ruy, Carnegie Mellon University, USA; Hajo Rijgersberg, Wageningen University & Research, The Netherlands; one anonymous reviewer

Abstract. Measurement units and their relations like conversions or quantity kinds play an important role in many applications. Thus, many ontologies covering this area have been developed. Consequently, for new projects aiming at reusing one of these ontologies, the process of evaluating them has become more and more time consuming and cumbersome. We evaluated eight well known ontologies for measurement units and the relevant parts of the Wikidata corpus. We automatically collected descriptive statements about these ontologies and used them for performing a comparison on an extensible collection of scripts. The computational results were finally merged, which uncovered several issues and misconceptions in the examined ontologies. The issues were reported to the ontology authors. This caused new bugfix releases in three cases.

In this paper we will present the evaluation results including statistics as well as an overview of detected issues. We thereby want to enable a well-founded decision upon the unit ontology to use. Further, we hope to prevent errors in the future by describing some pitfalls in ontology development – not limited to the domain of measurement units.

Keywords: Measurement unit ontology, ontology comparison, ontology evaluation, ontology quality

1. Introduction

Units of measurement are an essential part in many aspects of modern life: The correct handling of the scale a value is measured in is crucial not only in science, but also in trade, industry, and administration. A well documented use of units is especially important, when a project is carried out by different partners with different backgrounds. One of the most prominent examples of neglecting this fact is the crash of the Mars Climate Orbiter in 1999, which the NASA investigation board attributed to a mismatch of used units between two components of its software [13].

Similar integration challenges arise on an even larger scale in the context of Big Data: With the increasing need to integrate datasets of different origins, data annotations using Semantic Web technologies gain importance. Using a machine readable annotation is essential for (semi)automatic discovery, verification, and integration.

As part of these semantic descriptions and to cover the field of measurement units and related concepts, over the last years several projects were initiated to create respective ontologies [5,12,15,16].

Most of these attempts were embedded in bigger research projects and, thus, catered to their specific



doi: 10.3233/SW-180310



Choosing Good IRIs

Identity vs. Equality

Properties



Choosing Good IRIs

Identity vs. Equality

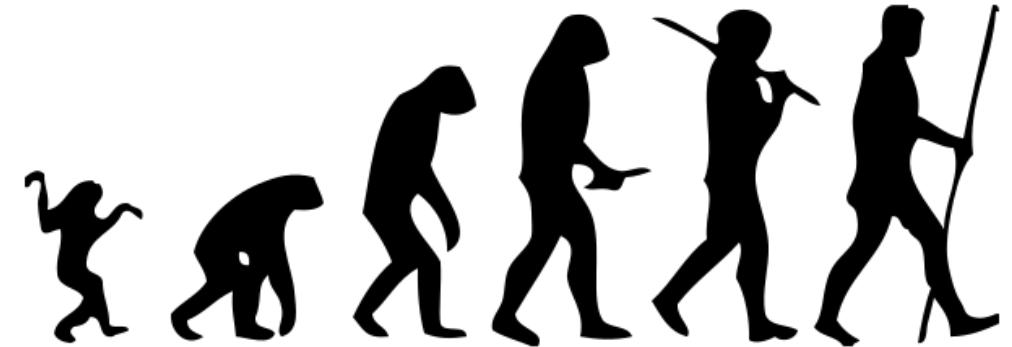
Properties



1a. IRIs should not contain the ontology version.

👎 Old versions may disappear from the web, breaking ...

- Resolving of IRIs
- owl:imports



ℹ Need references to specific versions? owl:versionIRI !

- Generic IRIs for reuse – always referring to latest version
- Version-specific ontology-IRIs for provenance, ...

http://www.wurvoc.org/vocabularies/om-1.6/Unit_of_measure

http://www.wurvoc.org/vocabularies/om-1.8/Unit_of_measure



1b. IRIs should not be too long.

👎 Cluttering (text-)editors

👎 Consuming valuable resources in restricted environments

- IoT RDF streams, ...

👍 Limit the length of your IRIs

👍 Generic local-names

- wd:Q67006310 “2nd International Workshop on Bad Or Good Ontology” (Wikidata.org)
- bfo:BFO_0000006 “spatial region” (BFO – OBOFoundry.org)
- ...



<http://www.ontology-of-units-of-measure.org/resource/om-2/constantCurrentThatProducesAnAttractiveForceOf2e-7NewtonPerMetreOfLengthBetweenTwoStraightParallelConductorsOfInfiniteLengthAndNegligibleCircularCrossSectionPlacedOneMetreApartInAVacuum>



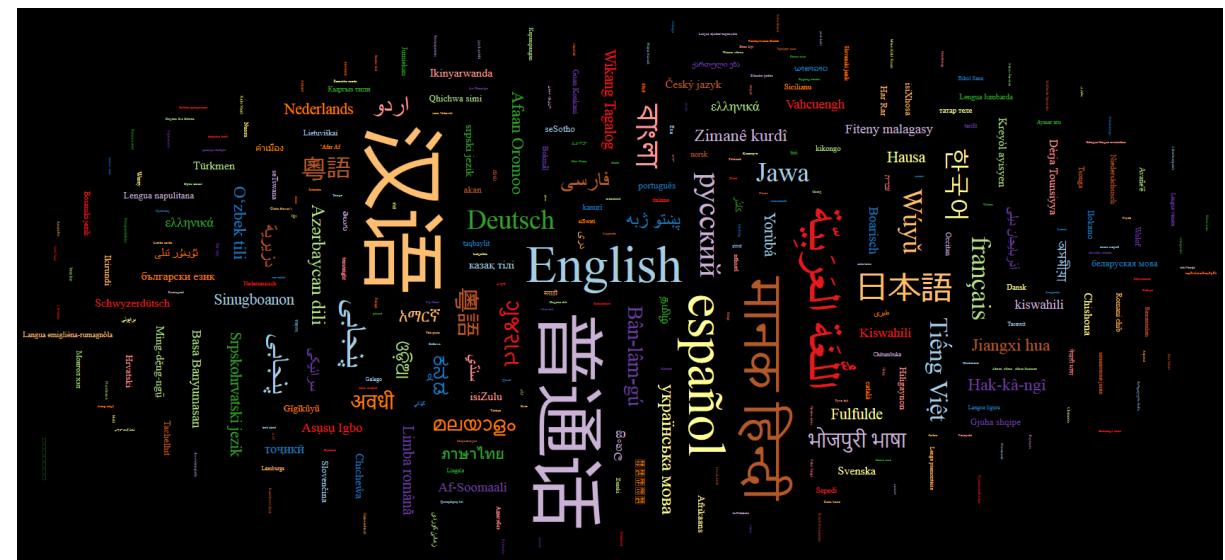
1c. IRIs of large resource collections should not contain natural language.

Fixing typos breaks references

Language dependent

Generic local-names and attached labels

- wd:Q13298 (Wikidata.org)
 - “Graz”@de
 - “Γκρατς”@el
 - “Graz”@en
 - “گراتس”@fa
 -



<http://purl.oclc.org/NET/muo/ucum/unit/pressure/pound-per-square-inch>

1d. Prefixes should not refer to multiple namespaces.

- In modularized ontologies

 Possible Namespace mix-ups

 Be consistent

 Use namespace lookups like <https://prefix.cc>



```
http://sweet.jpl.nasa.gov/2.3/reprSciComponent.owl
```

```
<!ENTITY comp "http://sweet.jpl.nasa.gov/2.3/reprSciComponent.owl#">
```

```
http://sweet.jpl.nasa.gov/2.3/statePhysical.owl
```

```
<!ENTITY comp "http://sweet.jpl.nasa.gov/2.3/matrCompound.owl#">
```

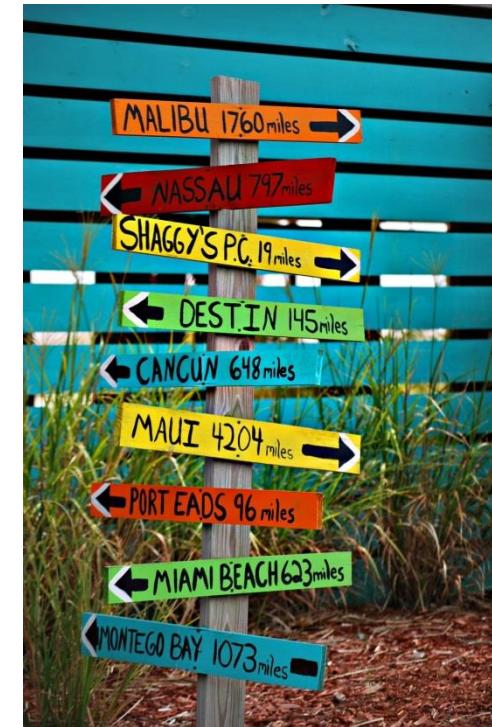
1e. Namespaces should not be referred by multiple prefixes.

- In modularized ontologies

👎 Namespace confusion

👍 Be consistent

👍 Check the results in generated ontologies



<http://sweet.jpl.nasa.gov/2.3/propOrdinal.owl>

<!ENTITY screla "http://sweet.jpl.nasa.gov/2.3/relaSci.owl#">

<http://sweet.jpl.nasa.gov/2.3/propEnergyFlux.owl>

<!ENTITY screla2 "http://sweet.jpl.nasa.gov/2.3/relaSci.owl#">

1f. Namespaces should not omit the hash.

👎 Fine with XML, but breaks in Turtle or SPARQL

- Hash # is used for comments

👎 Different prefixes/namespaces in reuse

```
PREFIX screla: <http://sweet.jpl.nasa.gov/2.3/relaSci.owl>
PREFIX units: <http://sweet.jpl.nasa.gov/2.3/reprSciUnits.owl>

SELECT ?quantity
WHERE {
  # get all quantities that use lumen
  ?quantity screla:#hasDefaultUnit units:#lumen .
}
```



<http://sweet.jpl.nasa.gov/2.3/propEnergyFlux.owl>

<!ENTITY screla "http://sweet.jpl.nasa.gov/2.3/relaSci.owl">



Choosing Good IRIs

Identity vs. Equality

Properties



2a. Do not confuse equivalency and identity.

2b. Be aware of alleged synonyms.

👎 Possible oversimplifications

👎 Mathematically equivalent – not identical

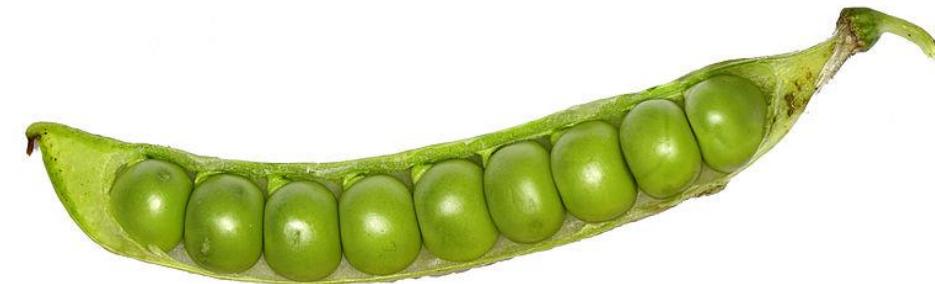
- Conversion factor of one

ℹ owl:sameAs means identical

- Both IRIs exchangeable in all contexts!
- Same for skos:altLabel etc.

ℹ Other levels of “sameness”

- E.g., skos:broadMatch, skos:narrowMatch,
skos:closeMatch, skos:exactMatch, ...



“cubic metre per square metre” vs “metre”

“litre” vs “cubic decimetre”

2c. Know the exceptions.

- Things are actually renamed sometimes.



- Linked by `owl:sameAs`



- Multiple labels, e.g., `skos:altLabel`



The gon (or grad, where grad is an alternative name for the gon) is an alternative unit of plane angle to the degree, defined as $(\pi/200)$ rad.

International Bureau of Weights and Measures (BIPM). The International System of Units (SI), 8 edition, 2014.



Choosing Good IRIs

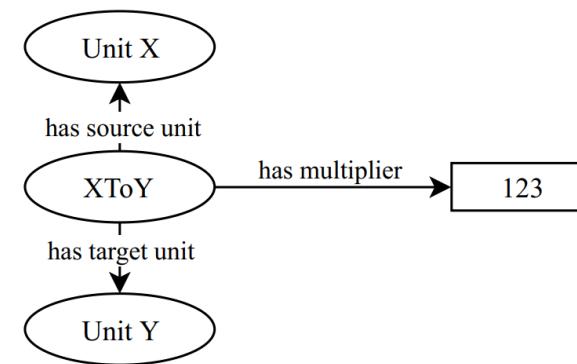
Identity vs. Equality

Properties

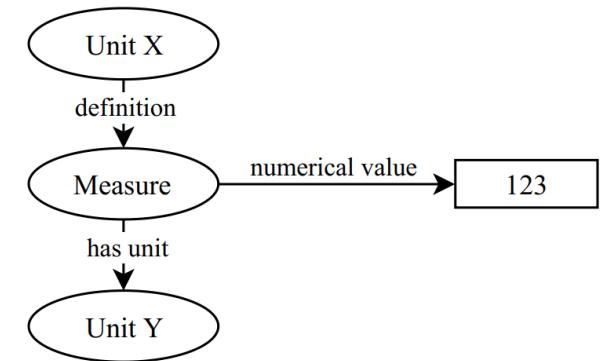


3a. Properties should be modeled resilient against misinterpretation.

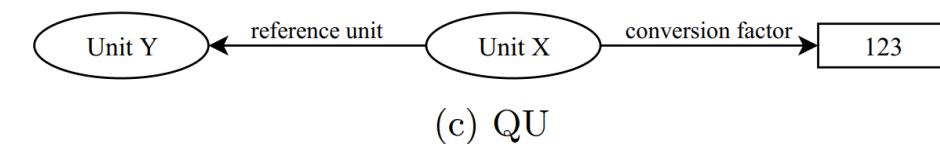
- Different models for unit conversions



(a) OBOE



(b) OM

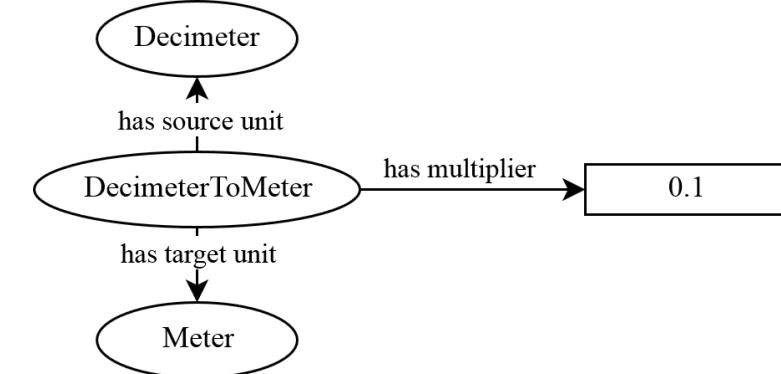


(c) QU



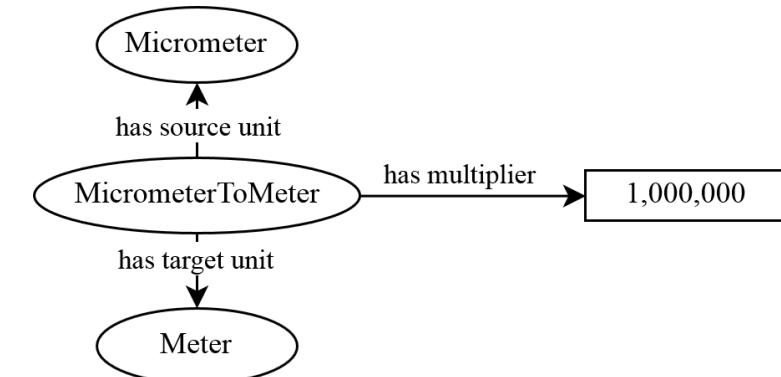
3a. Properties should be modeled resilient against misinterpretation.

- Different models for unit conversions



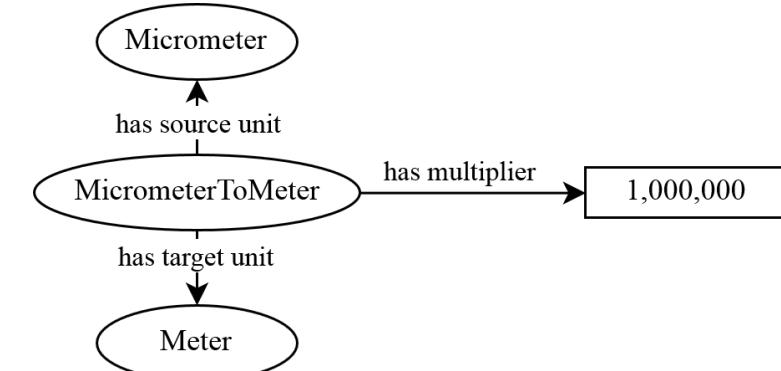
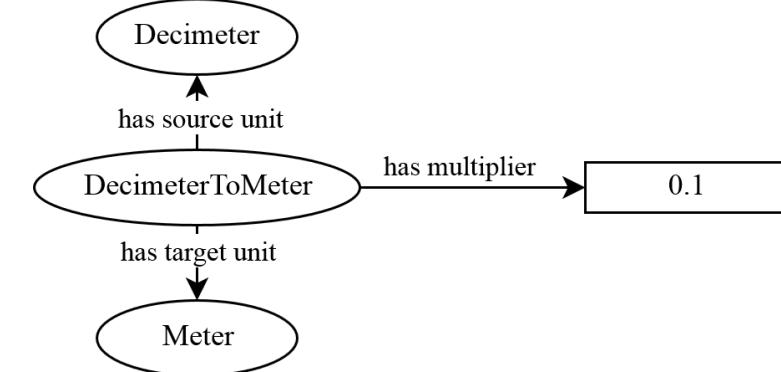
3a. Properties should be modeled resilient against misinterpretation.

- Different models for unit conversions



3a. Properties should be modeled resilient against misinterpretation.

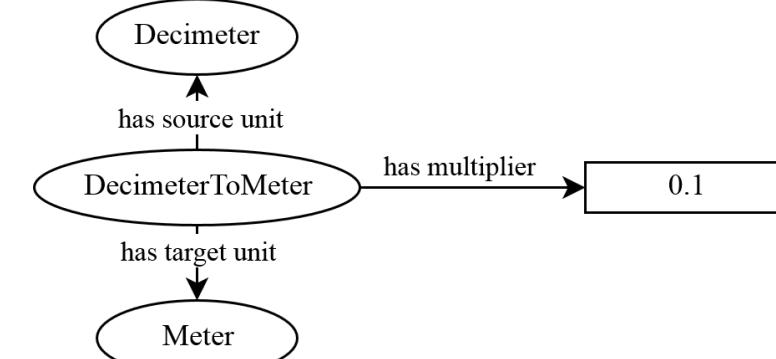
- Different models for unit conversions



3a. Properties should be modeled resilient against misinterpretation.

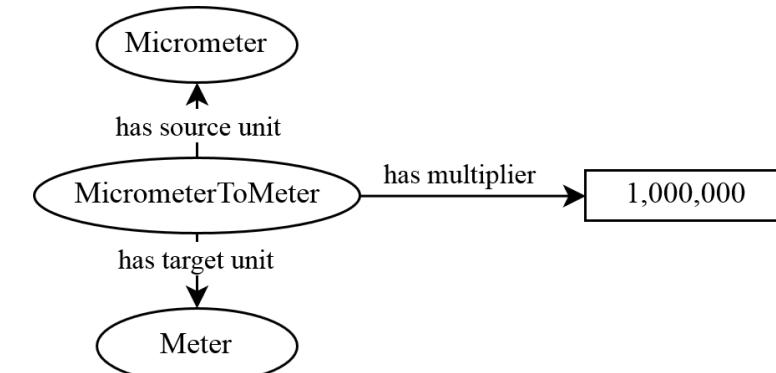
- Different models for unit conversions

👎 Unclear models



👍 Prevent misinterpretation

👍 Add descriptions



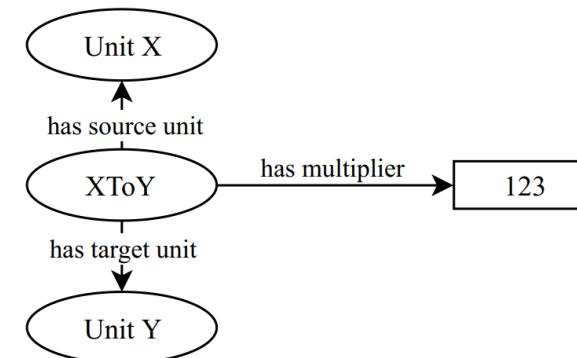
3a. Properties should be modeled resilient against misinterpretation.

- Different models for unit conversions

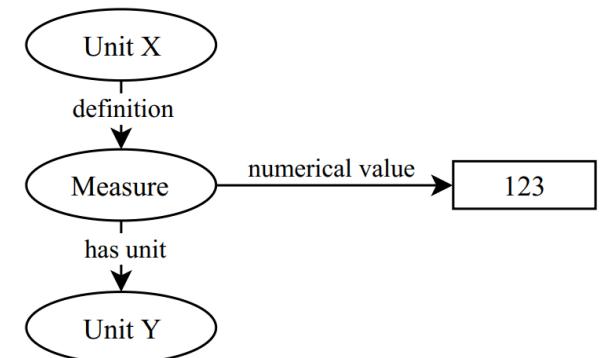
👎 Unclear models

👍 Prevent misinterpretation

👍 Add descriptions



(a) OBOE



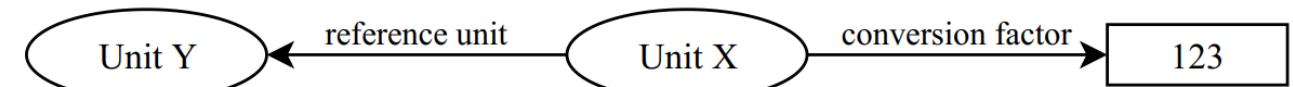
(b) OM



3b. Dependent properties should be encapsulated into distinct resources.

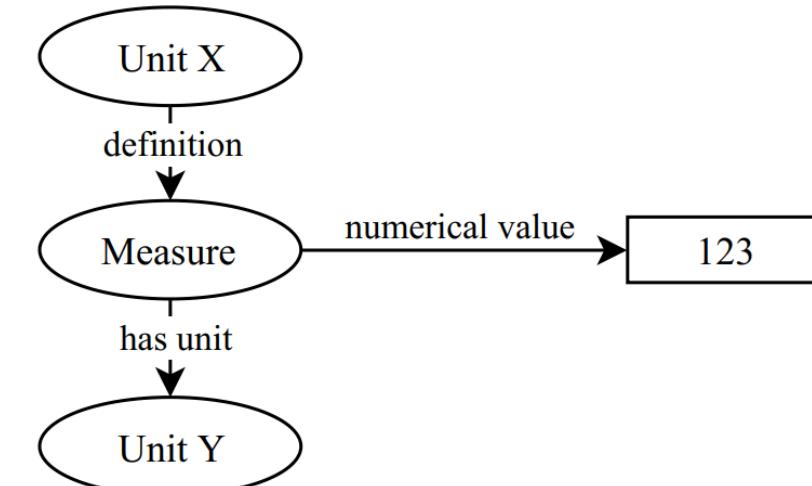
👎 Possible only once per resource

👎 Losing the dependency between values



👍 Add a distinct resource

- Preserve the dependency
- Allow for multiple definitions



Summary

👎 Issues encountered while reusing unit ontologies

- IRI design
- Identity vs. Equality
- Property Modeling

👍 Hints to prevent them



Besides tool support ...

- 👍 Reusability testing in ontology development
 - Usability testing in user interface design

- 👍 Have others use your ontology before publishing.



Building Ontologies for Reuse

Lessons Learned from Unit Ontologies

Sirko Schindler

Sirko.Schindler@dlr.de

Datamanagement-Technologies
Institute of Data Science
German Aerospace Center (DLR)
Jena, Germany



**Deutsches Zentrum
für Luft- und Raumfahrt**
German Aerospace Center

Jan Martin Keil

Jan-Martin.Keil@uni-jena.de

Heinz Nixdorf Chair for Distributed Information Systems
Friedrich Schiller University Jena
Jena, Germany



**FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA**



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA



Image Credits

- Motivation: https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg
- 1a: https://commons.wikimedia.org/wiki/File:Human_evolution_scheme.svg
- 1b: <https://pixabay.com/photos/letters-letter-wooden-alphabet-4466230/>
- 1c: data @ Wikidata
- 1d: <http://www.freestockphotos.biz/stockphoto/8213>
- 1e: <https://pxhere.com/en/photo/984500>
- 1f: <https://pixabay.com/illustrations/faulty-system-chair-broken-team-4295549/>
- 2a: https://commons.wikimedia.org/wiki/File:Green_pea_pod_8872.jpg
- 2b: https://commons.wikimedia.org/wiki/File:The_McClure_Twins.jpg
- 2c: <https://www.publicdomainpictures.net/en/view-image.php?image=11968&picture=unique-concept>
- Summary: <https://pixabay.com/illustrations/tropical-summer-sunset-beach-1651426/>
- Reusability testing: <https://knowyourmeme.com/memes/forwards-from-grandma>

