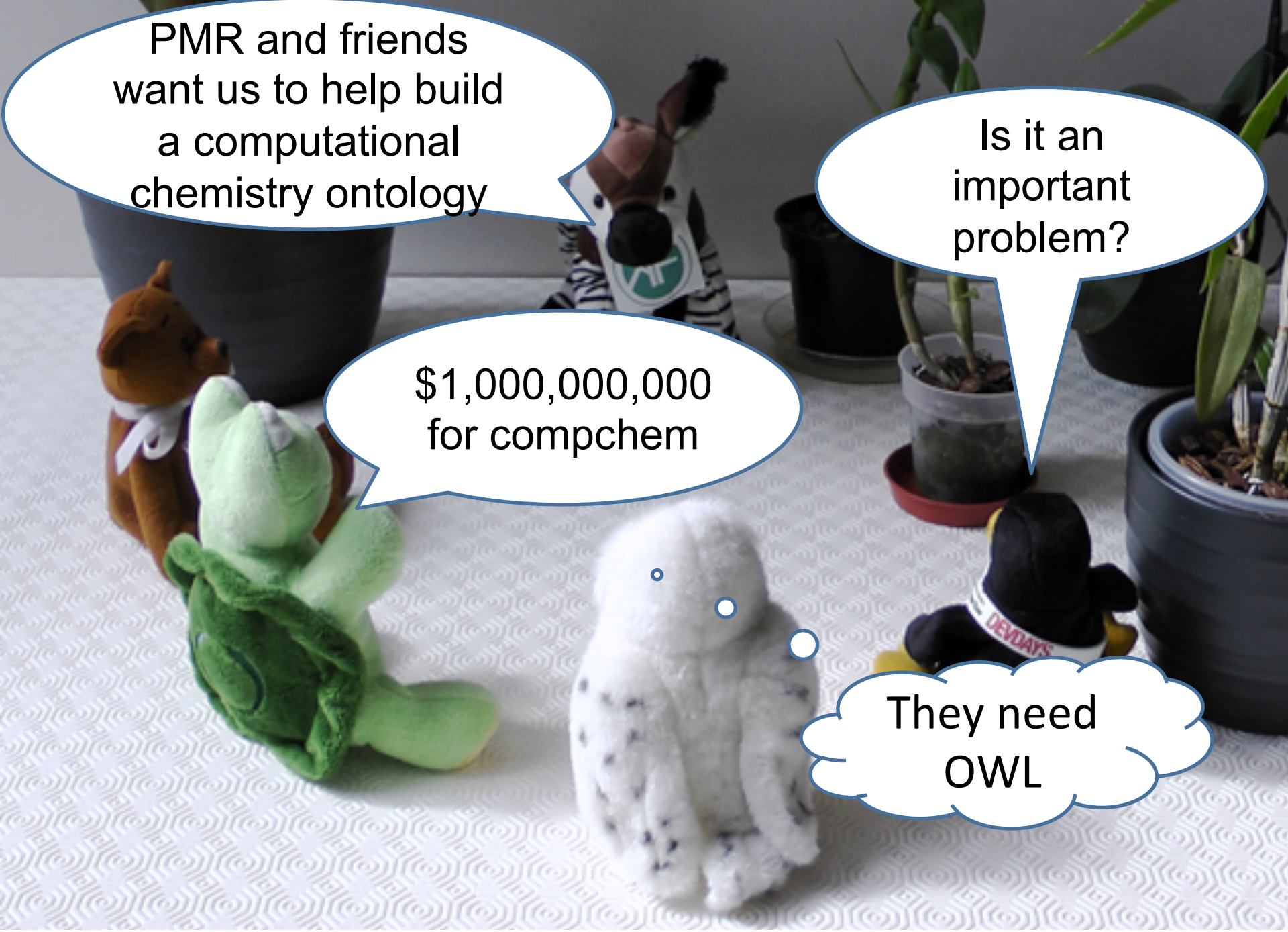
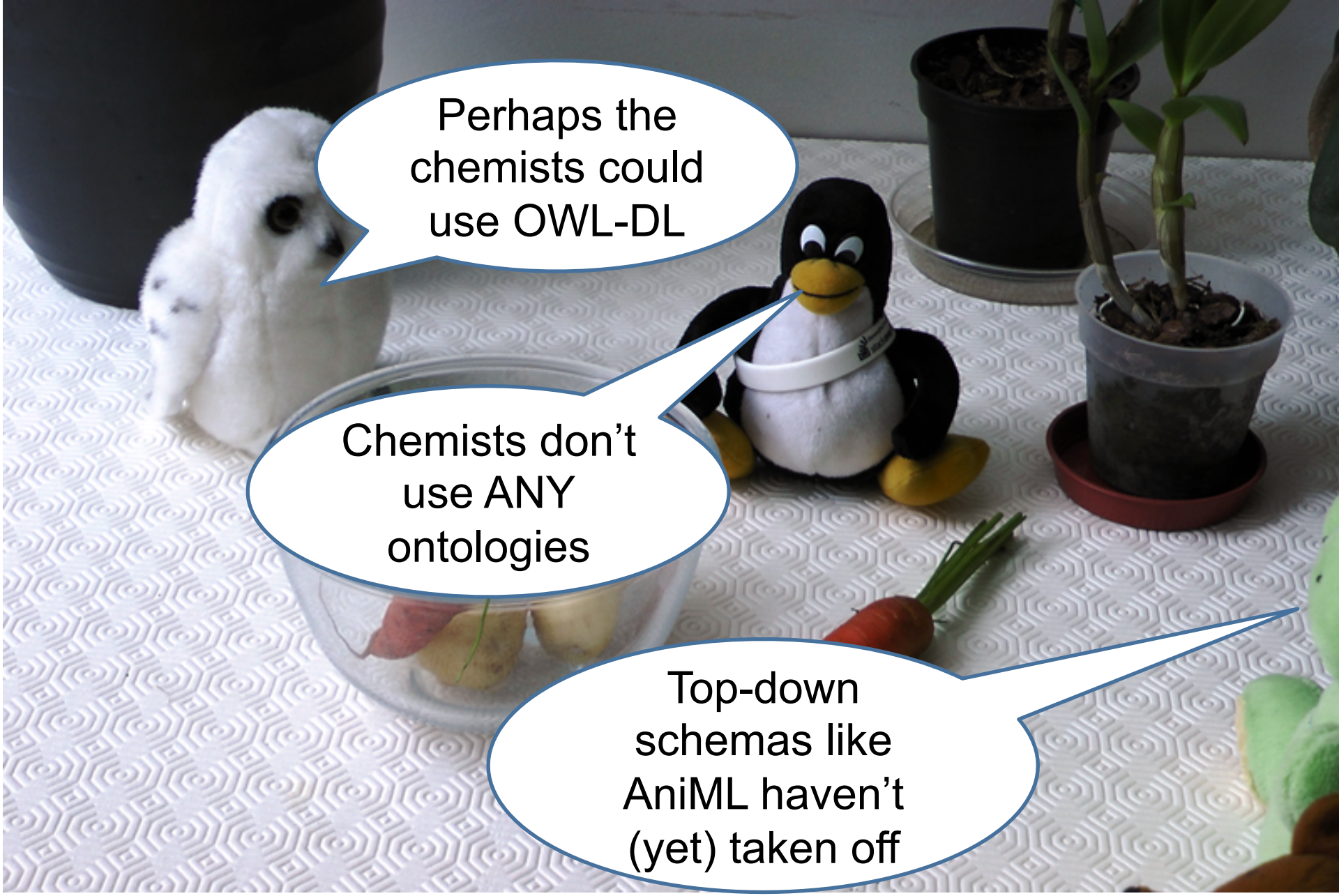
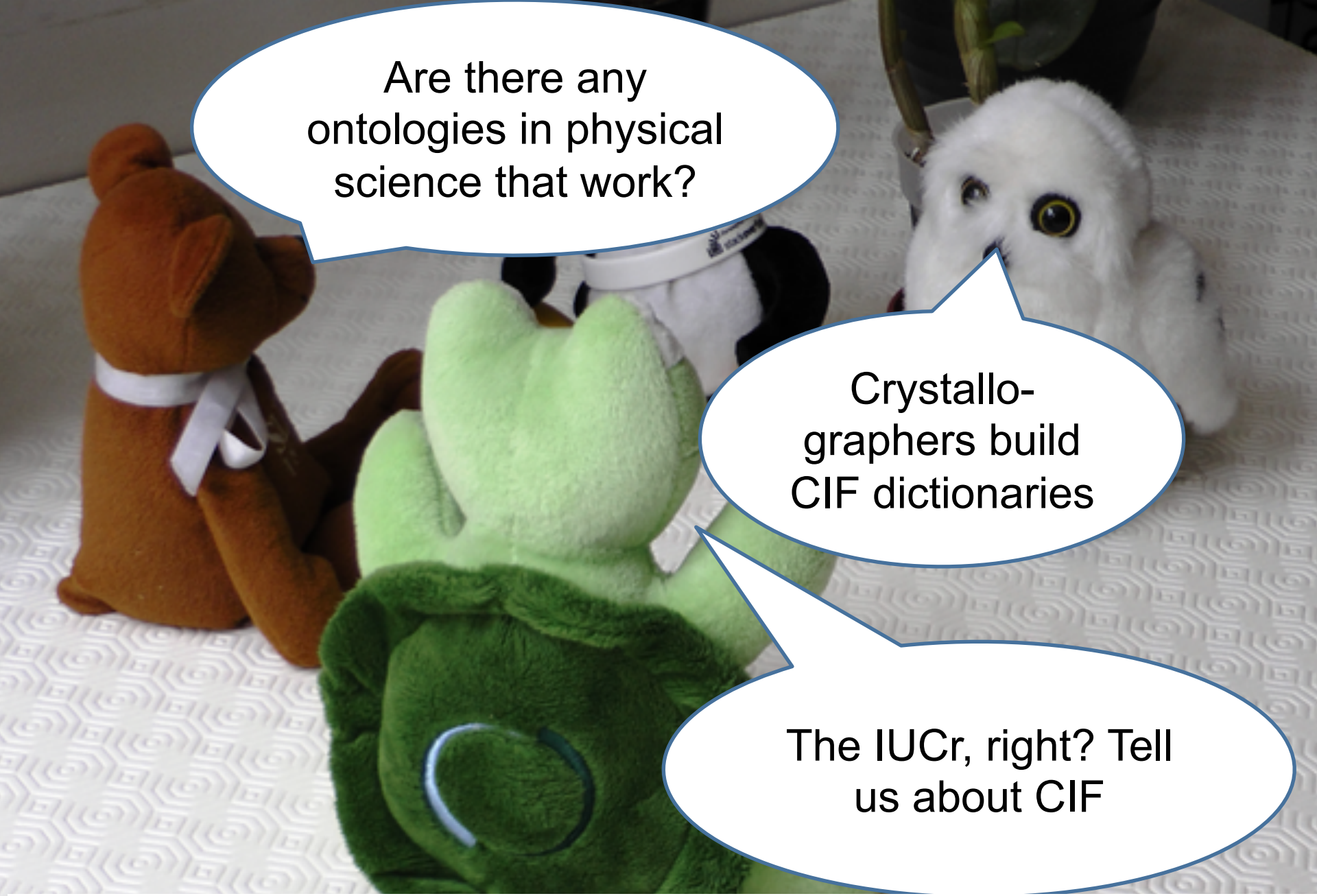
Ontologies in Physical Science

Onto Workshop, ed.ac.uk 2013-04-11

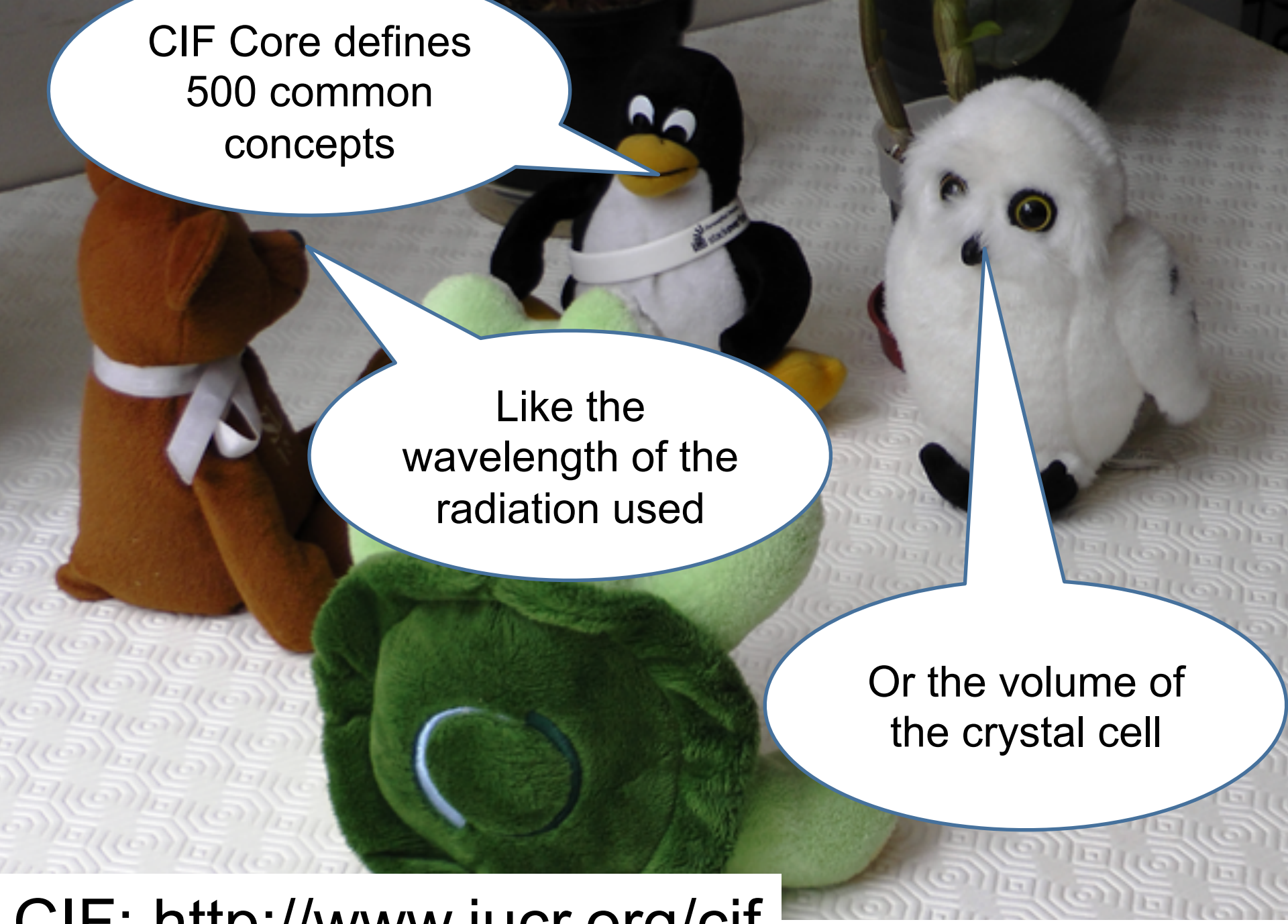
An #animalgarden production and
Peter Murray-Rust, OKF & cam.ac.uk


PMR and friends want us to help build a computational chemistry ontology
Is it an important problem?
$1,000,000,000 for compchem
They need OWL


Perhaps the chemists could use OWL-DL
Chemists don't use ANY ontologies
Top-down schemas like AniML haven't (yet) taken off



IUCr: International Union of Crystallography



CIF: http://www.iucr.org/cif

An example ?

Core dictionary (coreCIF) version 2.4.3

_diffrn_ambient_temperature

Definition: The mean temperature in kelvins at which the intensities were measured.

Range: 0.0 -> infinity **Type:** numb

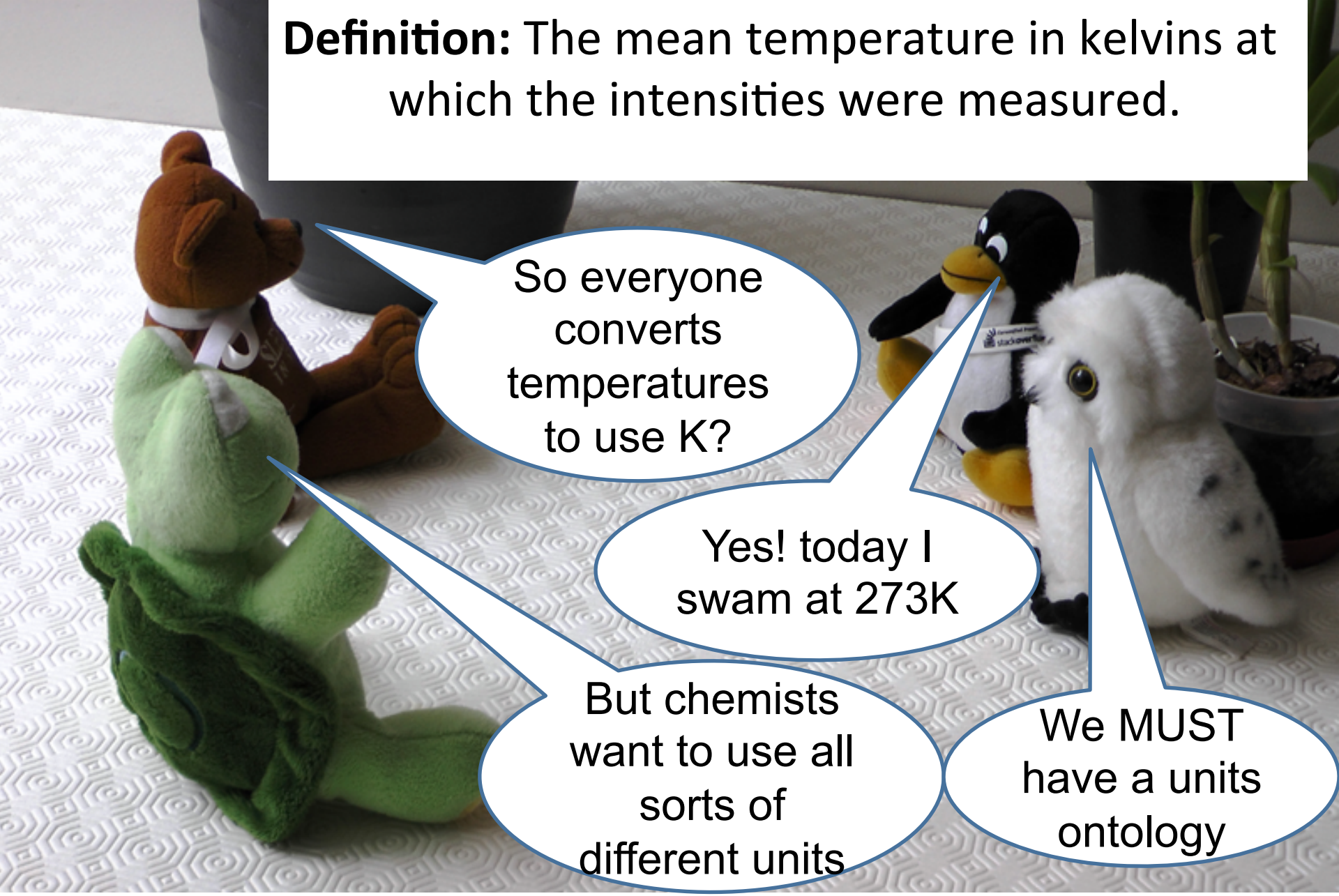
ID

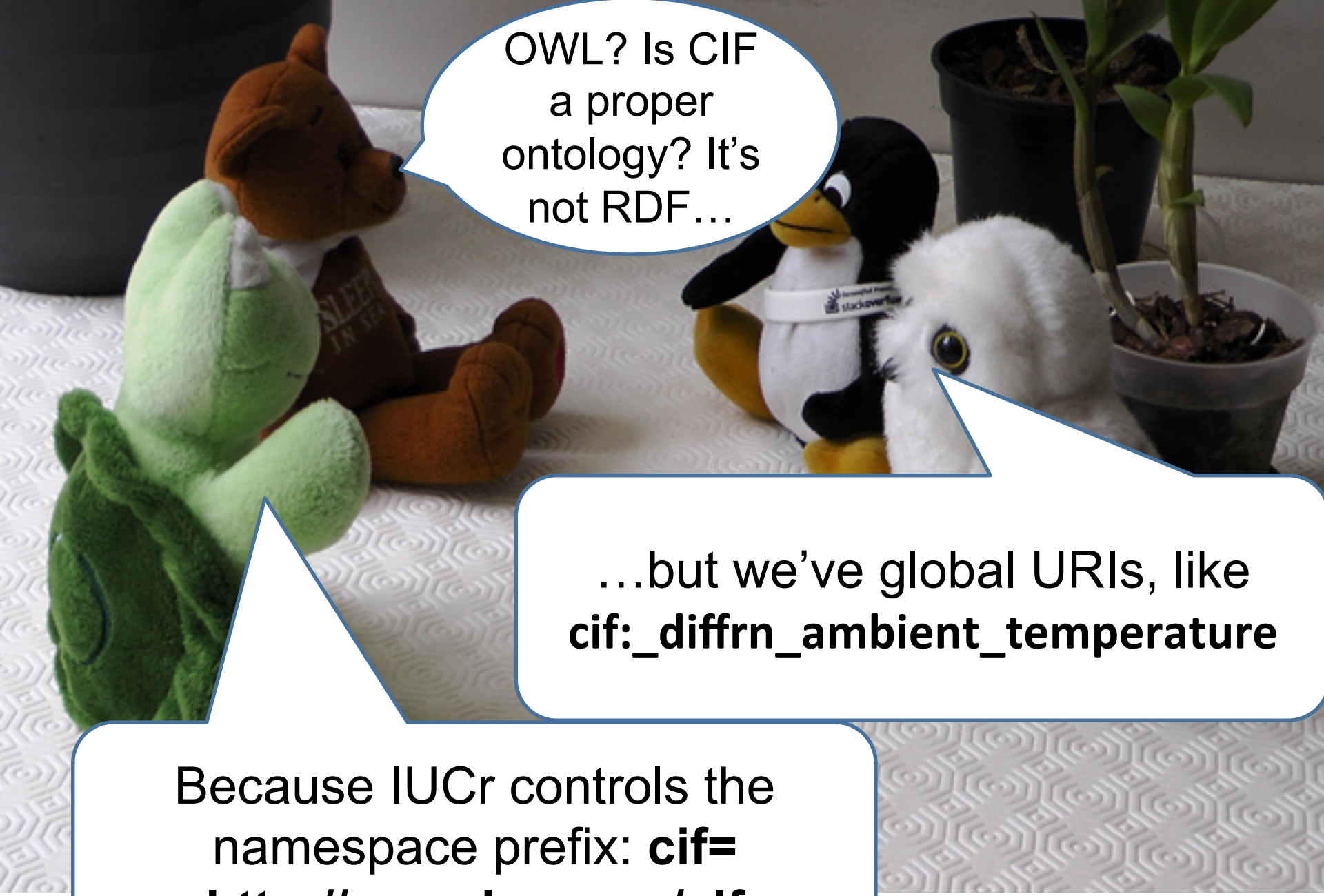
For machines: Constraint + type

For humans

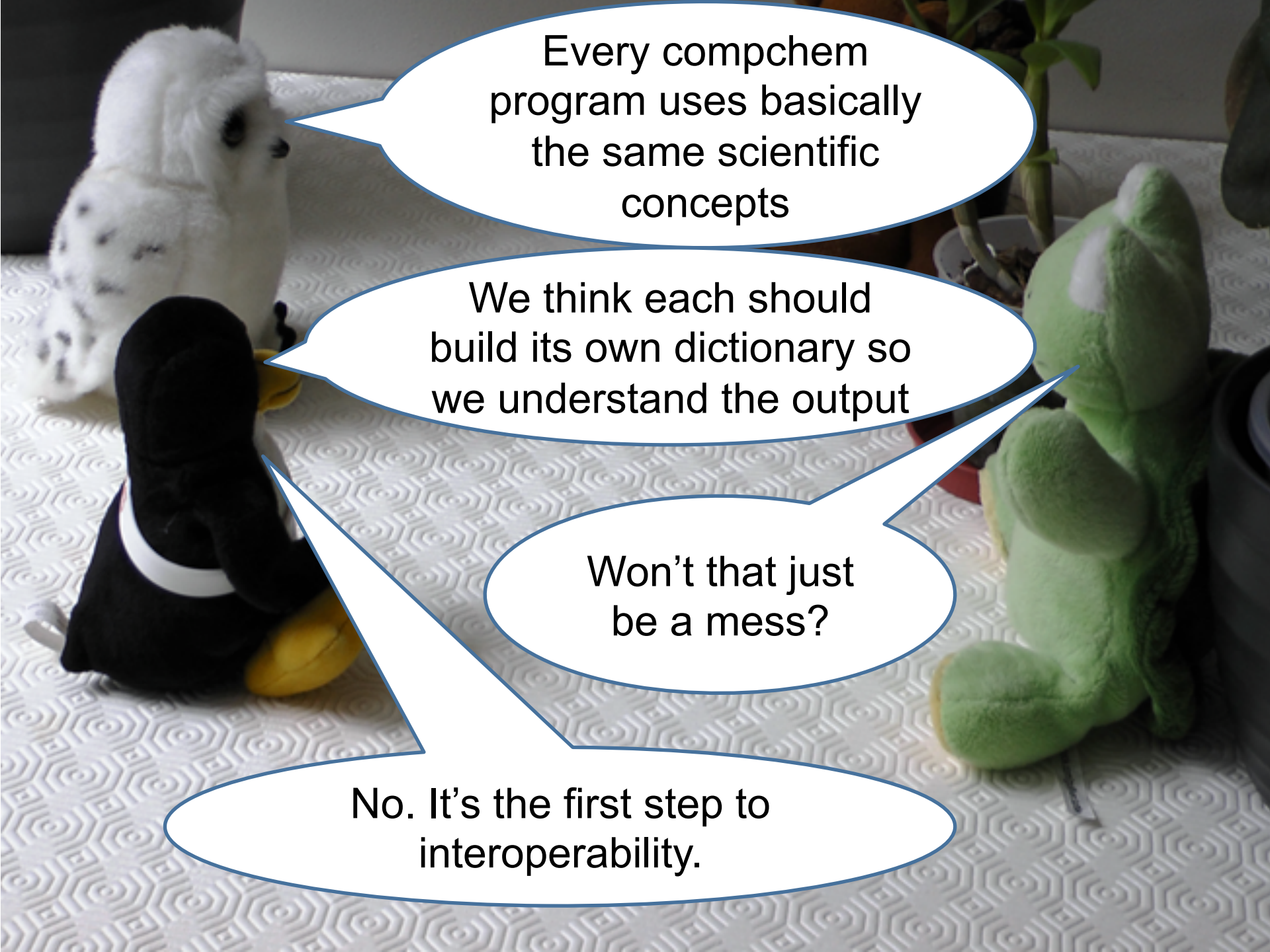
http://www.iucr.org/__data/iucr/cifdic_html/1/cif_core.dic/Idiffrn_ambient_temperature.html

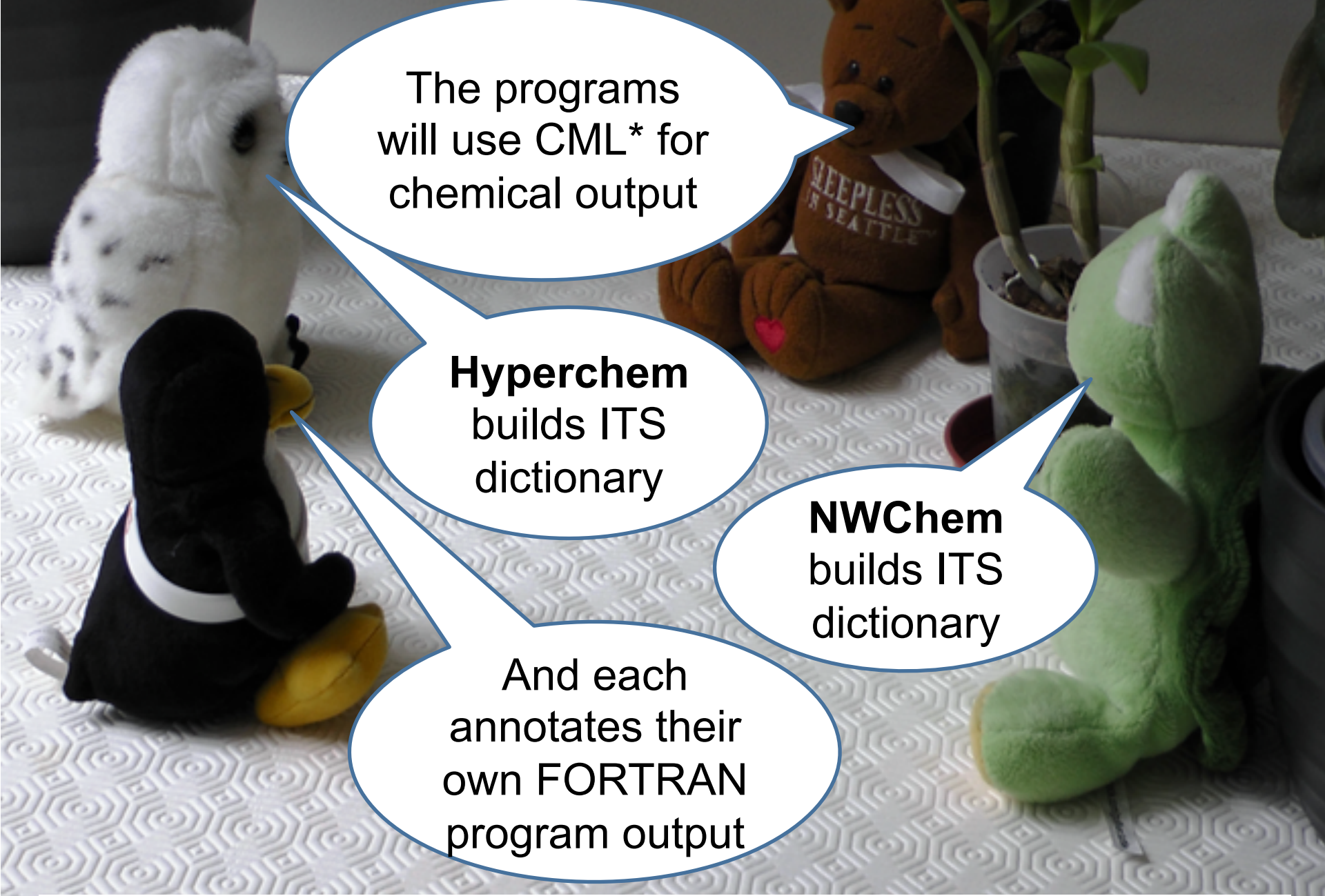
Definition: The mean temperature in kelvins at which the intensities were measured.



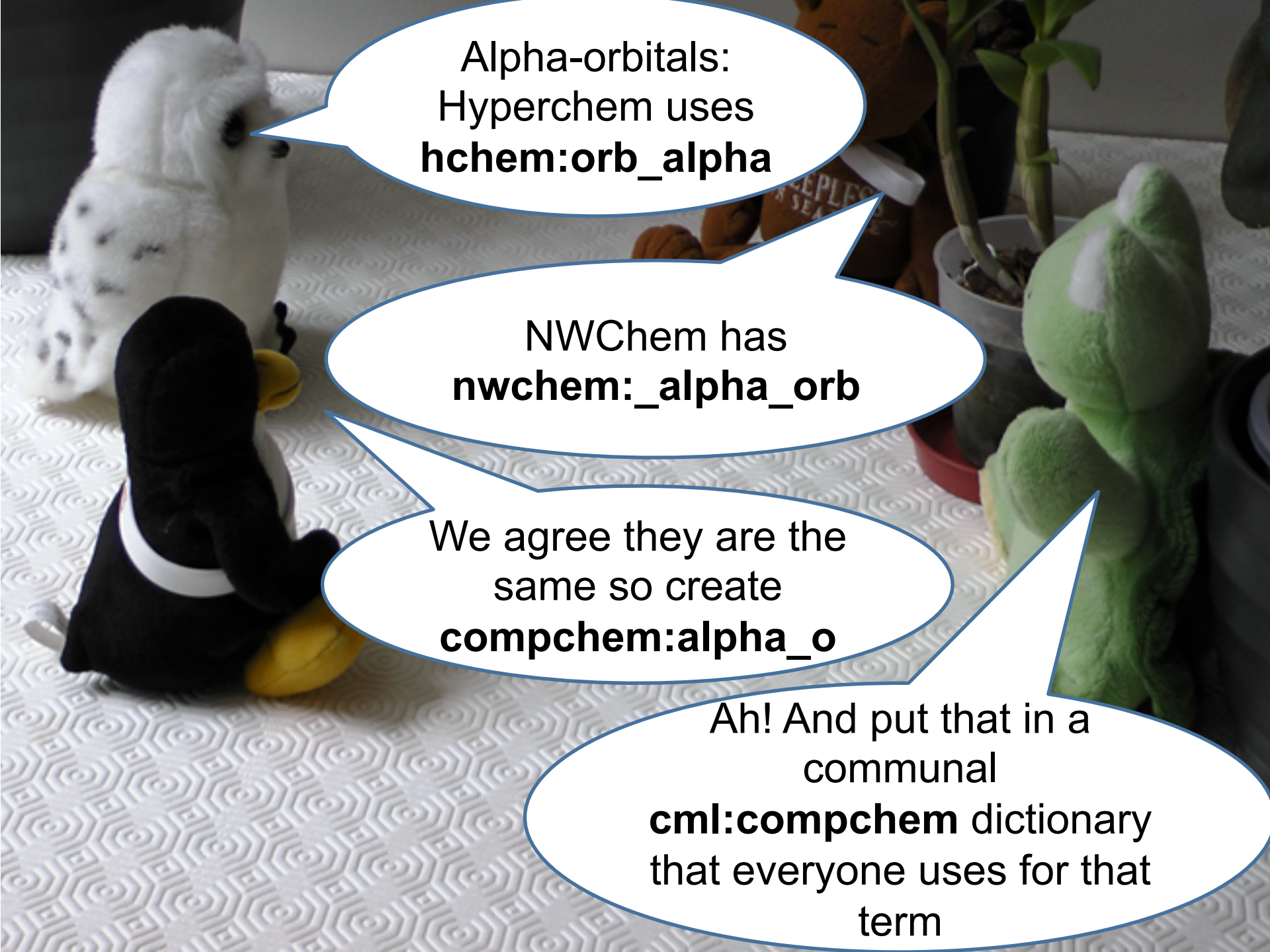

OWL? Is CIF a proper ontology? It's not RDF…
…but we've global URIs, like **cif:_diffrn_ambient_temperature**
Because IUCr controls the namespace prefix: **cif= http://www.iucr.org/cif**

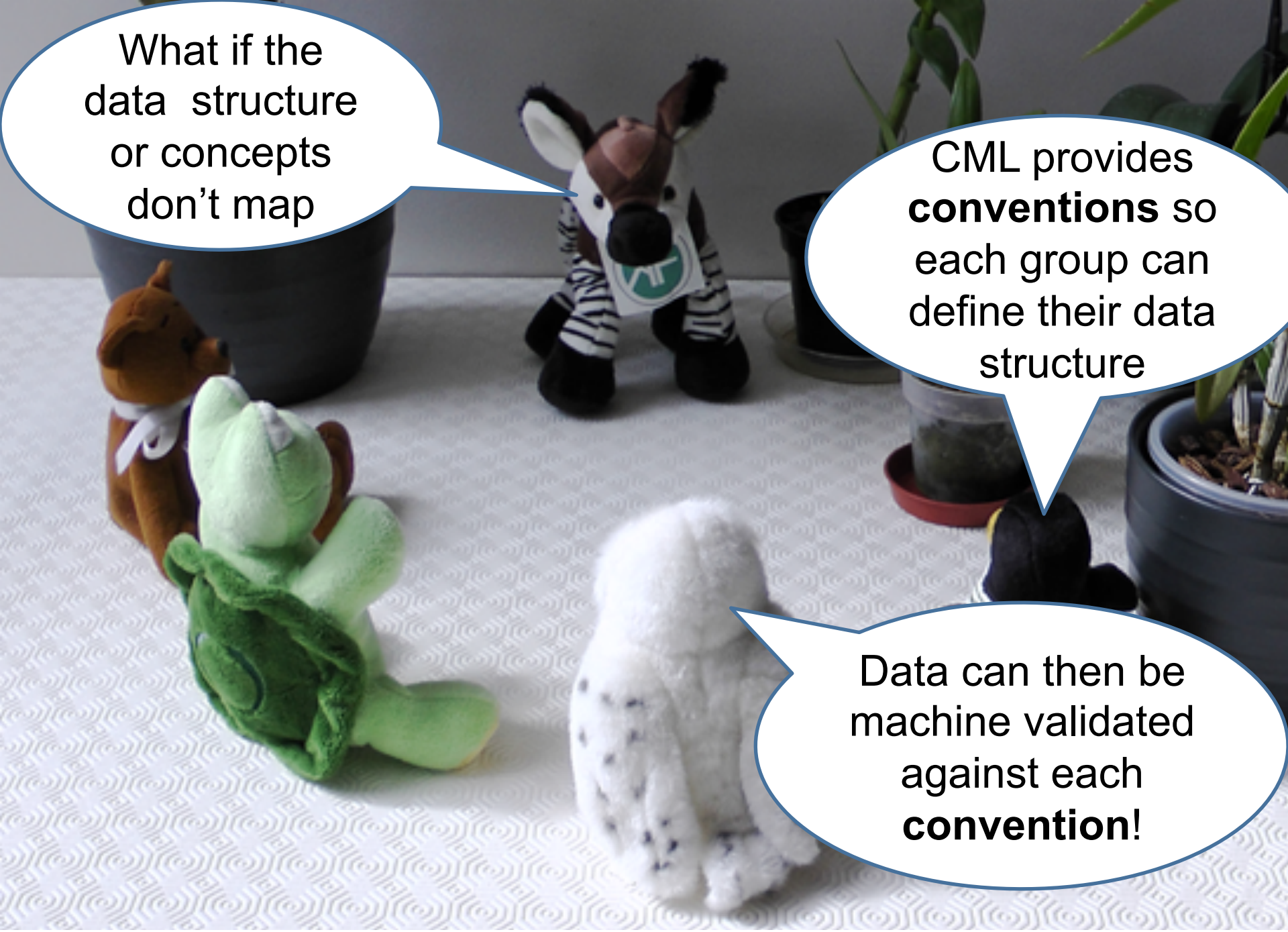

CIF had 20 years of community involvement through IUCr
But most top-down chemistry projects don't work
So we'll do this bottom-up.


Every compchem program uses basically the same scientific concepts
We think each should build its own dictionary so we understand the output
Won’t that just be a mess?
No. It’s the first step to interoperability.



* Chemical Markup Language http://www.xml-cml.org


Alpha-orbitals:
Hyperchem uses
hchem:orb_alpha
NWChem has
nwchem:_alpha_orb
We agree they are the
same so create
compchem:alpha_o
Ah! And put that in a
communal
cml:compchem dictionary
that everyone uses for that
term


What if the data structure or concepts don't map
CML provides **conventions** so each group can define their data structure
Data can then be machine validated against each **convention**!

TTT: Things Take Time (Piet Hein)


Will it work? It depends on people
National labs CSIRO/AU and PNNL/US are committed
I wish we had some publishers
And we have companies like Hyperchem and Kitware

Benefits of semantic dictionaries:

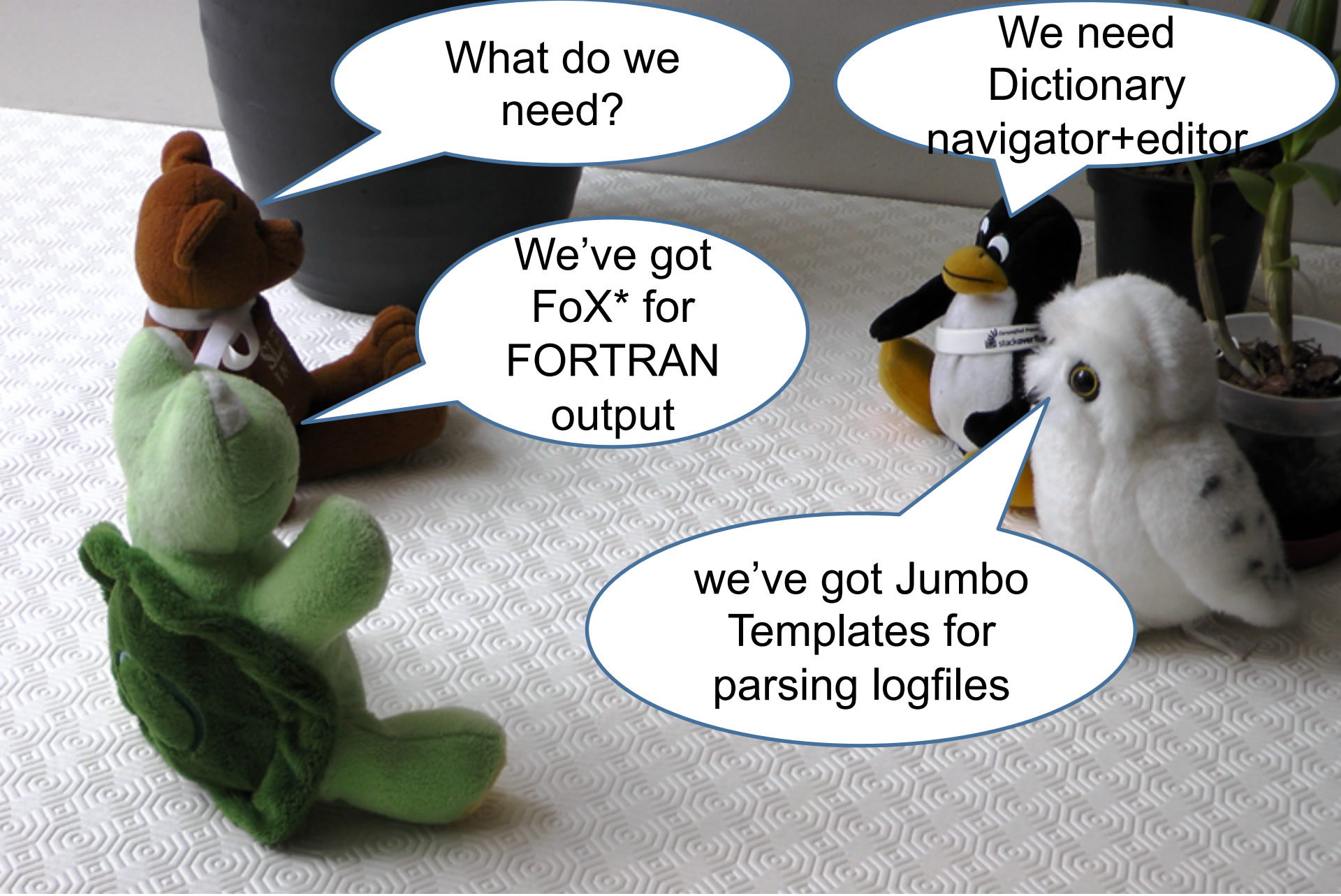
- FORTRAN logfile can be made semantic
- High degree of interoperability in chemistry
- Semantic publication (HTML5, CML, MathML)
- Interoperates with mainstream Web
- Easily scalable to other phys sci.

Problems:

- Closed code/minds is short-term market advantage
- Non-trivial commitment (updates, code revision)
- Getting top-down approval (e.g. IUPAC)


DEVDAYS


DEVDAYS


What do we need?
We need Dictionary navigator+editor
We've got FoX* for FORTRAN output
we've got Jumbo Templates for parsing logfiles