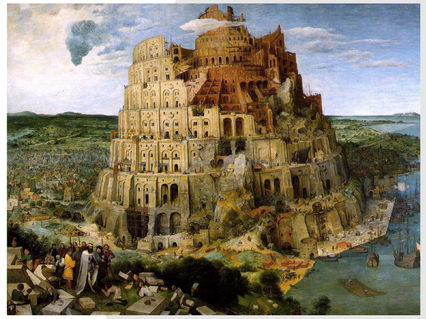
# WorldML vs. YaML – On the scope and purpose of mark-up languages

Jens Klump (1) and Robert Huber (2)
(1) German Research Centre for Geosciences, CeGIT, Potsdam, Germany (jens.klump@gfz-potsdam.de)
(2) MARUM, Univ. Bremen, Bremen, Germany (rhuber@pangaea.de)

EGU2011-10505
XL-249

## Motivation

With the rise of web services and XML came the idea that it should be possible to assemble components of a service oriented architecture in a modular fashion, similar to "plug and play" hardware. Information exchange was thought to be between machines only with no need for human intervention.



*The confusion of languages has troubled humanity for many centuries. The advent of machine-machine communication has highlighted the value of unabmbiguous communication.*

In practice, it turns our that in many instances communication between the parties operating the web services is still required. This is not a mere practical difficulty but is caused by an underlying property of communication that has been the subject of research and debate for centuries.

**Is there a way to predict which efforts to define a mark-up language or ontology will succeed and which will fail?**

**www.gfz-potsdam.de**
**www.marum.de**
**www.stratigraphy.net**



## WorldML or YaML?

In practice we can observe a widespread use of simple mark-up languages (e.g. Dublin Core), while more complex mark-up schemes (e.g. ABCDEFG) did not find wide spread application by their designated user communities.

Further complexities arise, if the XML schema allows application profiles deviating from the standard schema. Simple, purpose driven ontologies and "folksonomies" thrive, while top-down high level ontologies are still awaiting practical application.

## The Quest for a Perfect Language

There are principal considerations that can help us to decide where to direct our efforts in the design of mark-up languages and ontologies.

For centuries humans have dreamt of the perfect language that allows unambiguous communication. Philosophers have postulated that such a language we could even allow us to chart our knowledge and define its boundaries.

Philosophical work in the early 20th Century, such as Ludwig Wittgenstein's theory on language, investigated whether it is possible to define a "precision language" that would allow communication without ambiguity. However, further work on theory and experiments showed that language is a social construct where terms and meaning are defined through negotiation between communicating parties.

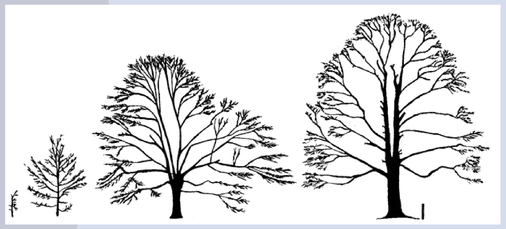
This finding has been supported by further work in systems theory (e.g. Luhmann) and computer linguistics (e.g. Solé, Baronchelli).

## Fork Ahead!

The fundamental cause for incertainties in these kind of information models are the bifurcations encountered in the course of defining information concepts. Bifurcations are most obvious in tree structures, but are also present in networks.

The characteristics of bifurcations can be illustrated by taking the image of a caterpillar moving from a leaf of a tree to the roots. The path from the leaf to the root is easy to define. Not so the other way round, from the root to a particular leaf.

Similarly, it is easy to define the connections within a given set of concepts, while a high level, top down approach has no obvious target concepts.

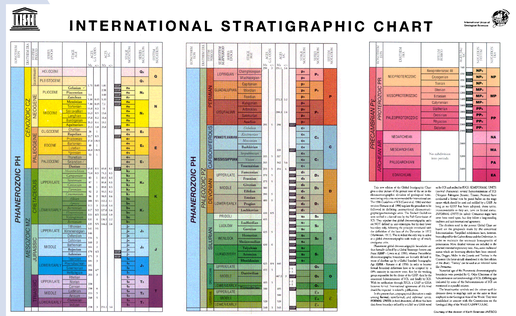


*The path from any leaf to the root of the tree is easy to find. The path from the root to a particular leaf is much more difficult.*

And yet, even if the "leaves" (concepts) are a finite, known set there are still infinite ways to relate concepts to eachother (Luhmann). Thus, even if the concepts are given, their semantic relationships may still be a matter of negociation.

## Conclusions

These fundamental considerations can be applied to the design of mark-up languages and ontologies. They show that there are limits to generalisation and, in turn, give an important role to negotiation in the definition of language elements and their meaning. This is particularly important in the distinction between top-down and bottom-up definition of mark-up languages and ontologies.




*Stratigraphy is an example of a bottom-up ontology. The set of concepts - even though it is large - is finite. The semantic relationships between concepts are well defined.*

In summary, ontology projects can be classified into two types:

**Type 1:** The set of concepts is potentially infinite and constructed top down. This type of project is likely to fail. If the set of concepts is limited by editorial considerations, the project automatically converts into Type 2.

**Type 2:** The set of concepts is finite and constructed bottom up. This type of project is likely to succeed, it is mainly an editorial task.

Keep it simple and targeted!