OGC
Open Geospatial Consortium

Metadata Reloaded

A meta level approach to

Metadata

...the real Meat in Data

Metadata DWG

Arnulf Christl
http://www.metaspatial.net

Content

- What is Metadata
- Design and Thing
- Catalogs
- INSPIRE and digital Geo Data
- New Requirement: Pragmatics
- ...and the w-holy REST

What is Metadata?

- [...]
- Metadata is data.
- However, it is impossible to identify metadata just by looking at it.
- We don't know when data is metadata or just data.

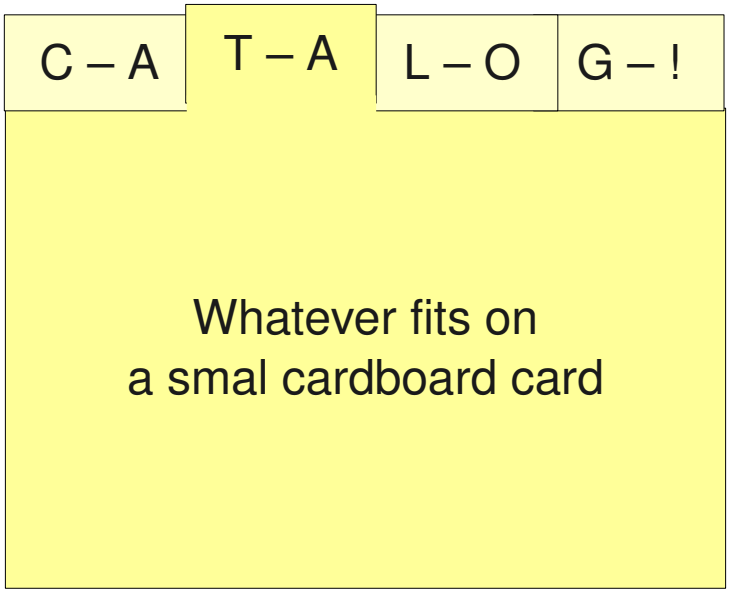
http://en.wikipedia.org/wiki/Metadata

OGC

Meta Data in Analog Catalogs...

The typical meta data of a book (the material object) can be:

- the name of the author
- the edition
- the year of publication
- the publisher
- and the ISBN number …

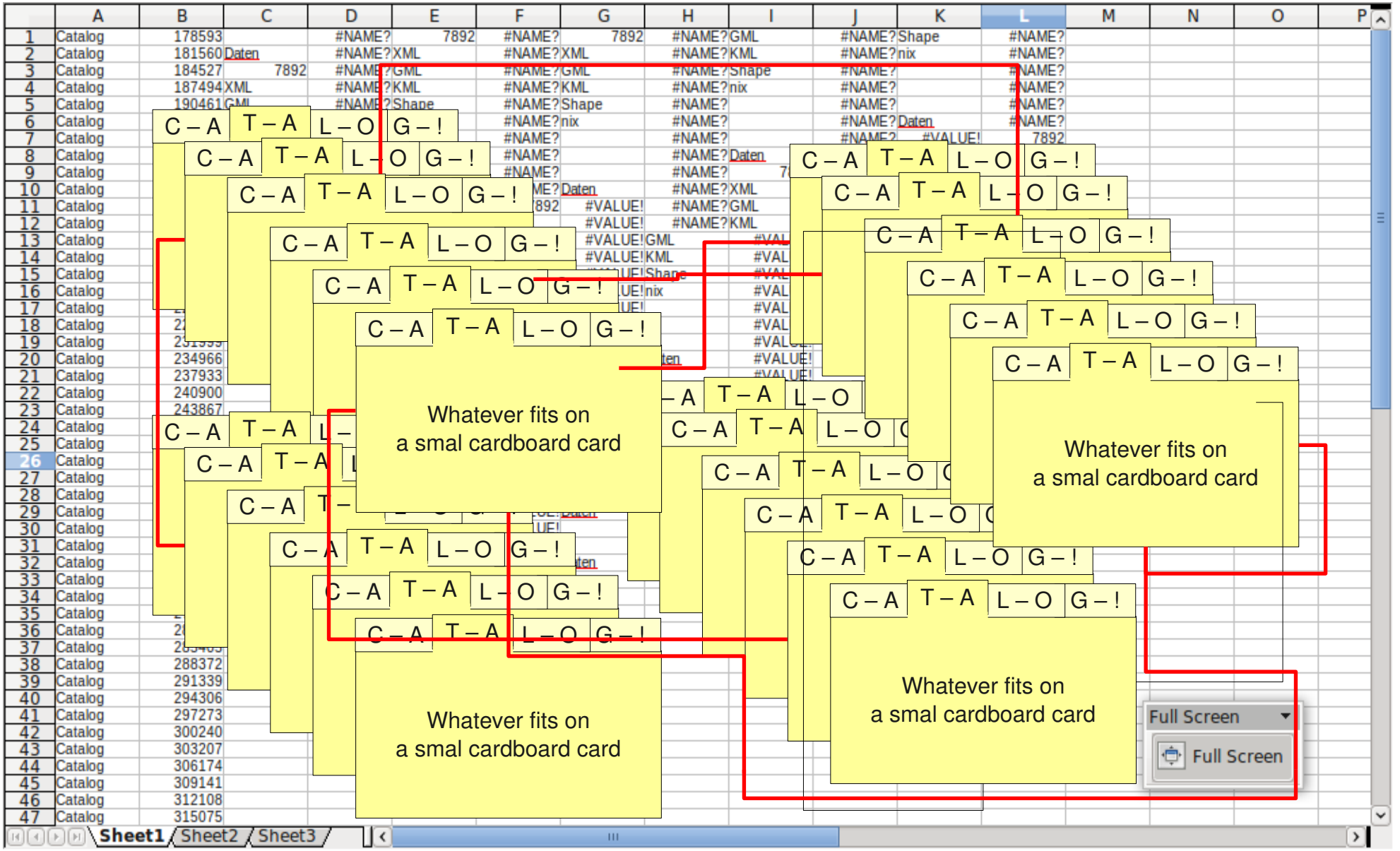


OGC

Therefore

it may be natural to think that digital meta data will look like a complex catalog

>



OGC

But...

digital
electronic
(Meta)–Data

...is a very different beast

OGC

Just for the fun of UTF-8
let's call this:

Metædata

"e" as in digitally electronic

OGC

Design vs. Thing

This presentation, information, any data in electronic format is nothing but design based on our concepts

Excursion I

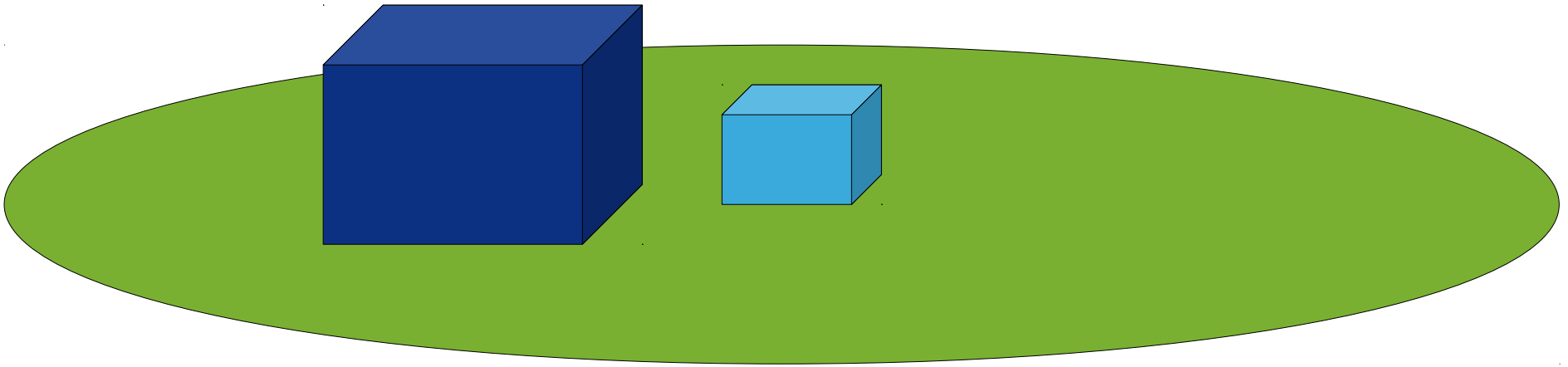
OGC

The World of Things

Things
take up space:

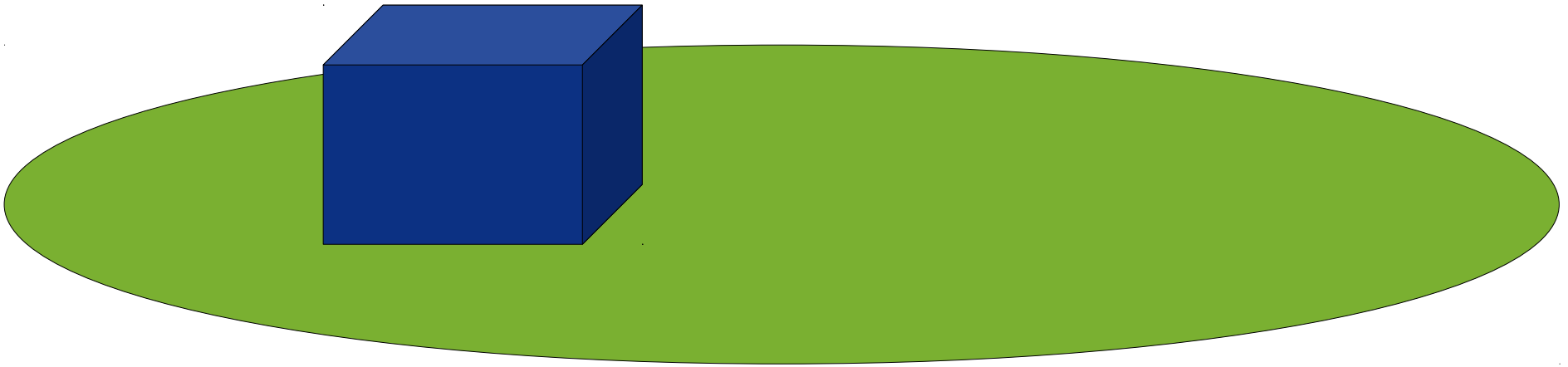
Matter is compounded of atoms and molecules and bound
by gravity and
order

Matter (a thing) can only be in one location at one time.



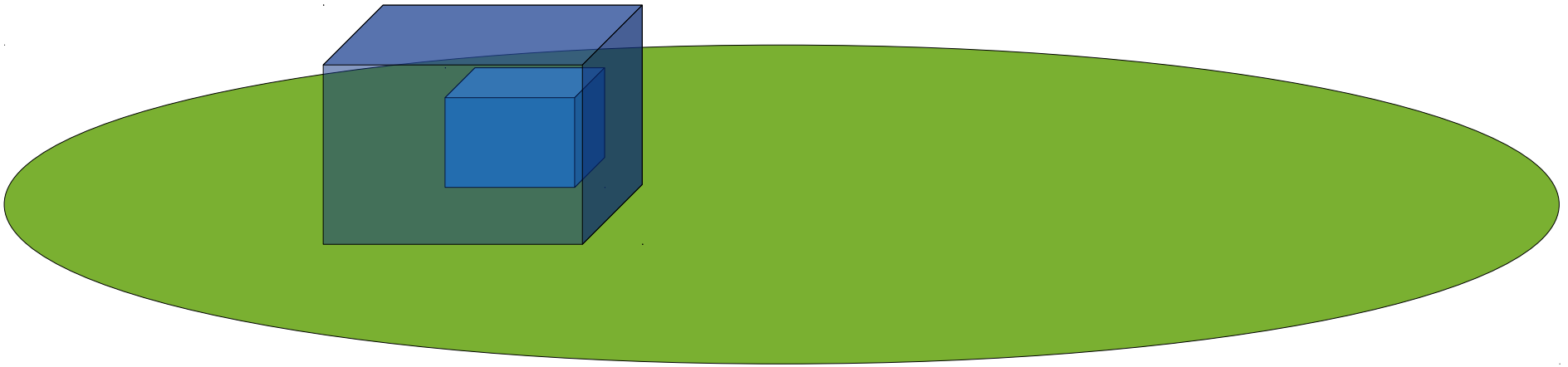
OGC

Things are always ordered and can thus hide each other.



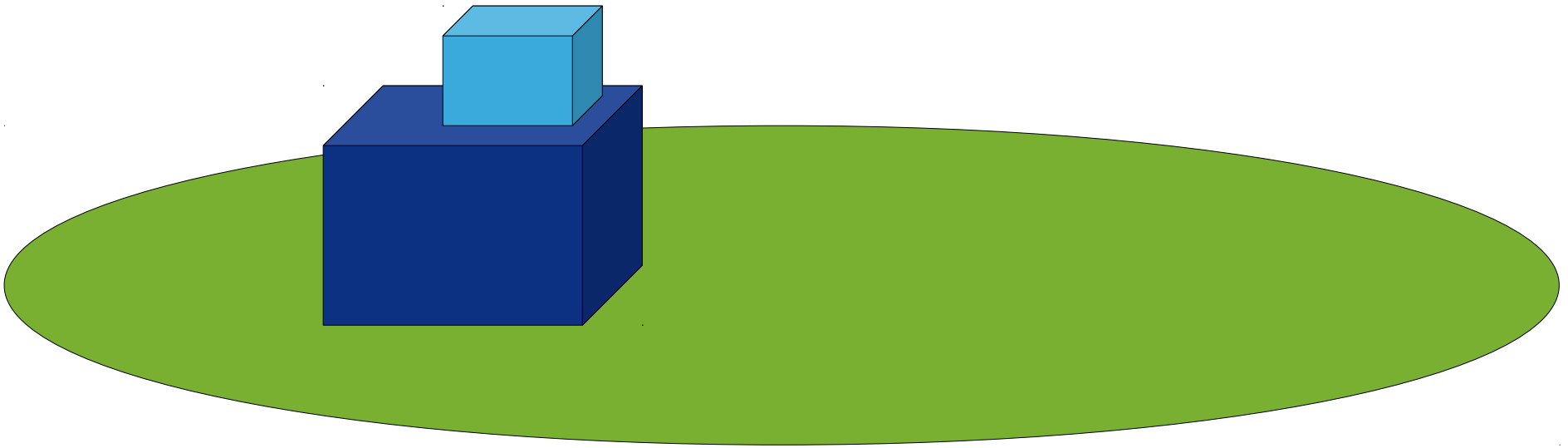
OGC

Just in case you didn't believe...



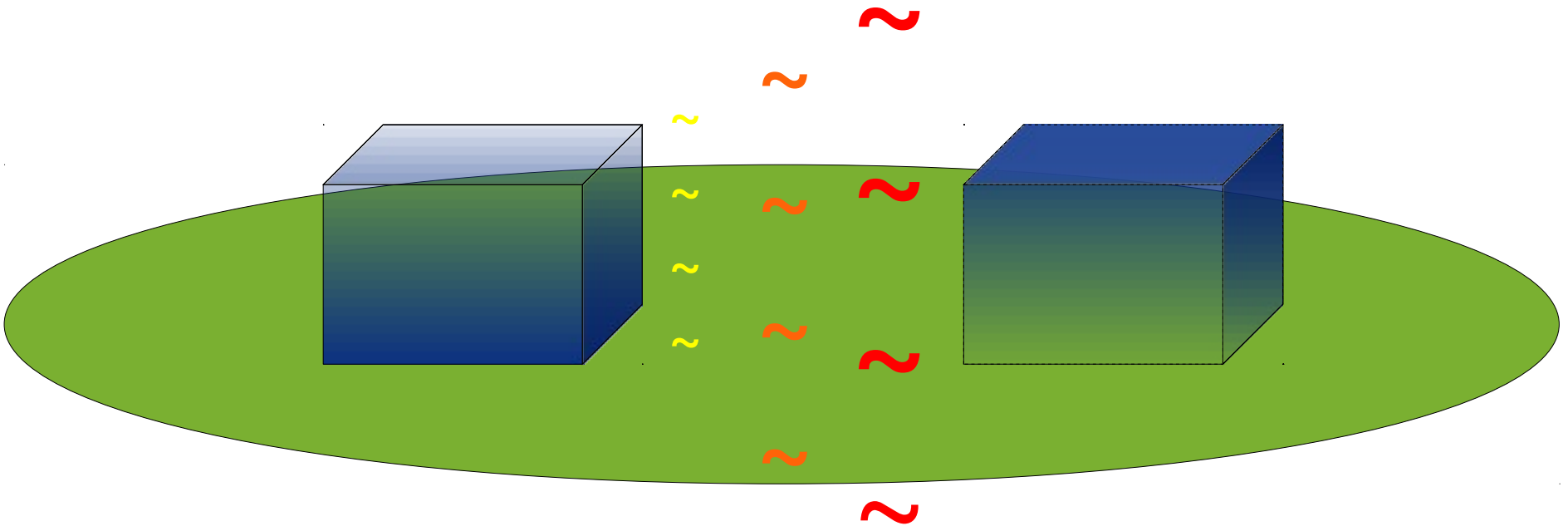
OGC

Two things cannot be at the same location at the same time.

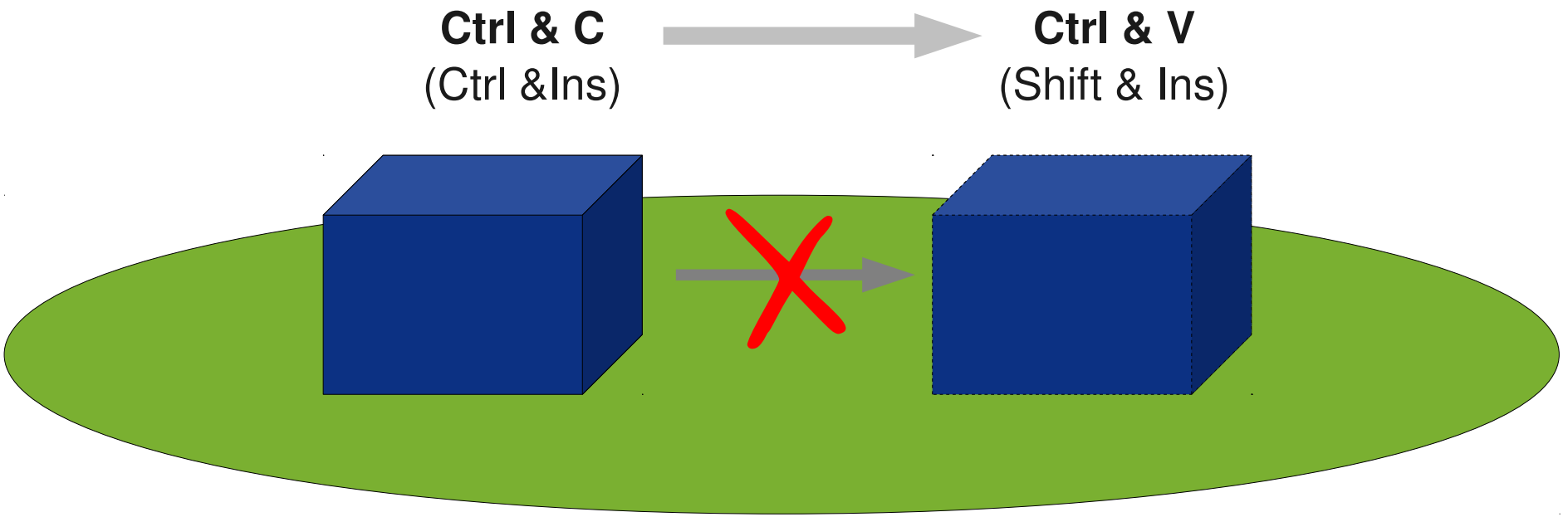


OGC

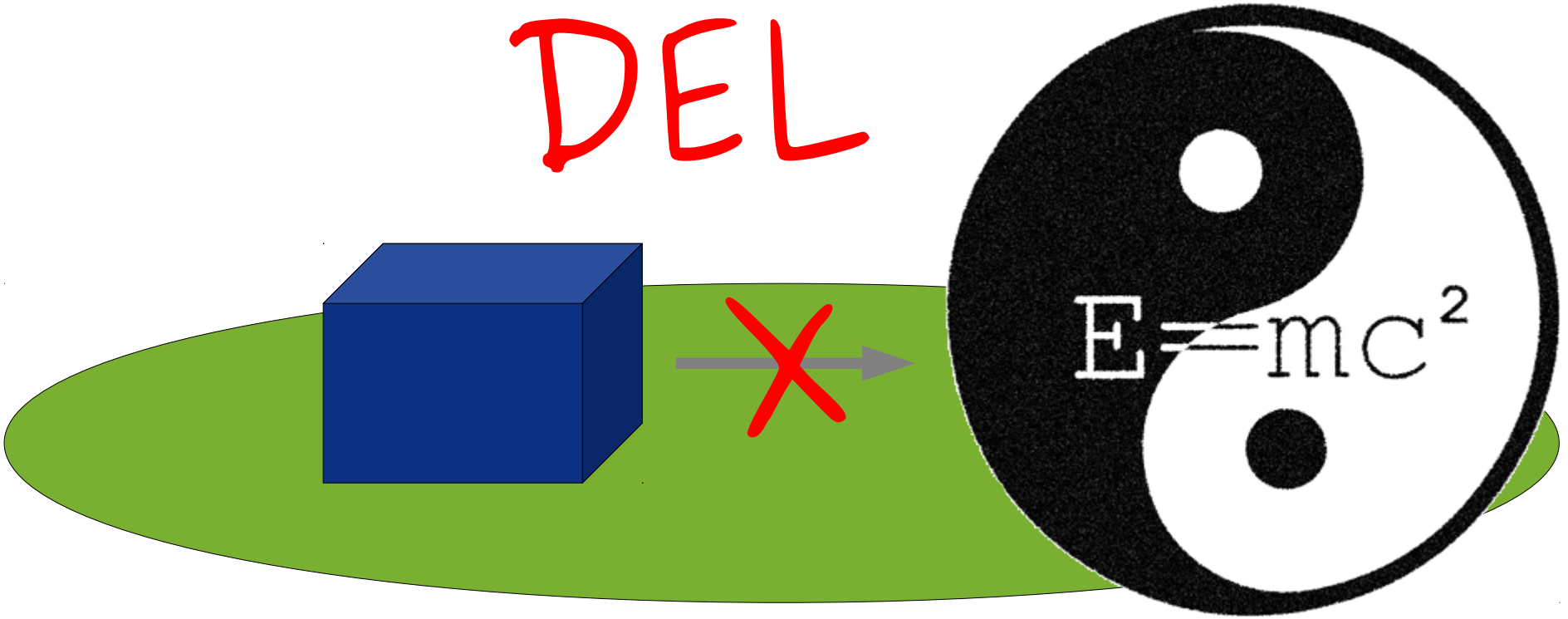
One thing cannot be at two locations at the same time.



Things cannot be copied!



Things cannot be deleted!



http://commons.wikimedia.org/wiki/File:E=mc2.png

Summarizing Matter

- Matter cannot exist without space.
- Matter takes up space.
- Matter can only be in one place at a time.
- Matter can not be copied.
- Matter can not be deleted.
- Matter can not be **linked**.

you know where this is leading...

OGC

Data can!

*...and Metædata is even **meant** to...*

OGC

…be copied,
*modified, **linked** and deleted.*

Excursion II

Copying is not theft!

OGC

Digital metædata can be at any place at any time

Metædata can be duplicated, copied, modified, deleted and **linked**!

... a bit.ly Open Source...

OGC

...digital Metadata?

Examples of computer data file metadata:

- The file name
- Access permissions
- The data of last access
- The file format (odt, shp, kml, xml, rss, rdf)
- And so on...

...Geospatial Meta æ Data

can contain and **is** metædata in itself:

kml - Keyhole Markup Language (OGC)

xml - eXtensible Markup Language (W3C)

rss - Really Simple Syndication (W3C)

rdf - Resource Description Framework (W3C)

File extensions are nothing but Metædata:
implicitly linked to format ***specifications***

INSPIRE'd Metadata?

"Profiles! Why, but why?"

...you can hear them complain.

The reason is simple enough, the idea is from the last millennium.

This is not to say that ideas are bad because they are old!

But they should be allowed to follow Life's § 1:

Evolve!

OGC

INSPIRE

The definition of "Metadata" in the INSPIRE directive is rather thin.

[…]

6. ‘metadata’ means information describing spatial data sets and spatial data services and making it possible to discover, inventory and use them;

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF

INSPIRE

2. Metadata shall include information on the following:

[...]

(b) conditions applying to access to, and use of, spatial data sets and services and, where applicable, corresponding fees;

[...]

(e) limitations on public access and the reasons for such limitations, in accordance with Article 13.

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF

Excursion II

2. Metadata shall include information on the following:

[...]

(b) conditions applying to access to, and use of, spatial data sets and services and, where applicable, corresponding fees;

[...]

(e) limitations on public access and the reasons for such limitations, in accordance with Article 13.

End-User License Agreement

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF

OGC

Metadata in INSPIRE

INSPIRE structures data on the top level with a rough category similar to the Dewey Decimal Classification.

The Annexes of the EU-Directive define what **type** of datasets need to be described and the ISO Standard 19139 defines **how** to describe them.

(There is no definition of what meta data is)

Example 1

The metadata for an orthorectified aerial photography image could contain:

- Spatial extent
- Coordinate reference system
- Projection
- Format, access options
- Date of exposure
- Resolution of original image
- Number of Bands
- Recording device
  - Digital
  - Analog
- Processing steps
  - Rectification
  - Geo referencing
  - Contrast adjustment
  - Brightness adjustment
  - etc.

Example 2

The metadata of a Traffic Information Service can consist of:

- Spatial extent
- Coordinate reference system
- Projection
- Format, access options
- Source of geometric data
- Actuality
- Acquisition method
  - Official data
  - Vounteered data
  - Forecasting method
- Consideration of construction sites, events, speedometers,
- etc.

Syntax and Semantics

Syntax and semantics are disciplines of the science of Semiotics.

In geoinformatics **syntax** describes spatial data **formally**, whereas **semantics** describes it's **meaning**.

To be able to store this information in a catalog it has to be indexed and structured.

Thesaurus and Indexing

A thesaurus is a networked collection of controlled vocabulary terms. A thesaurus uses:

- Associative relationships
- Parent-child relationships.
- The expressiveness of the relationships can vary

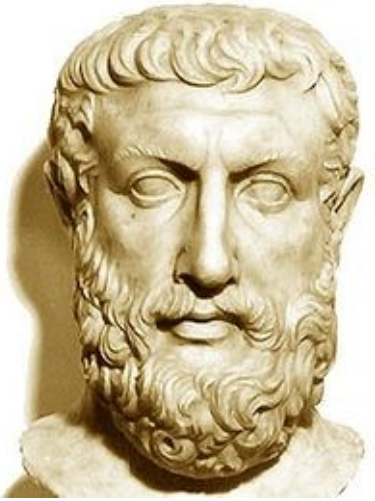
A common understanding of the **Meaning** is a prerequisite.

OGC

Ontology

Ontology is the philosophical study of the **nature of being**, existence or reality in general, as well as the basic categories of being and their relations.

Relations



Parmenides

http://twitter.com/sevenspatial/status/19854195289

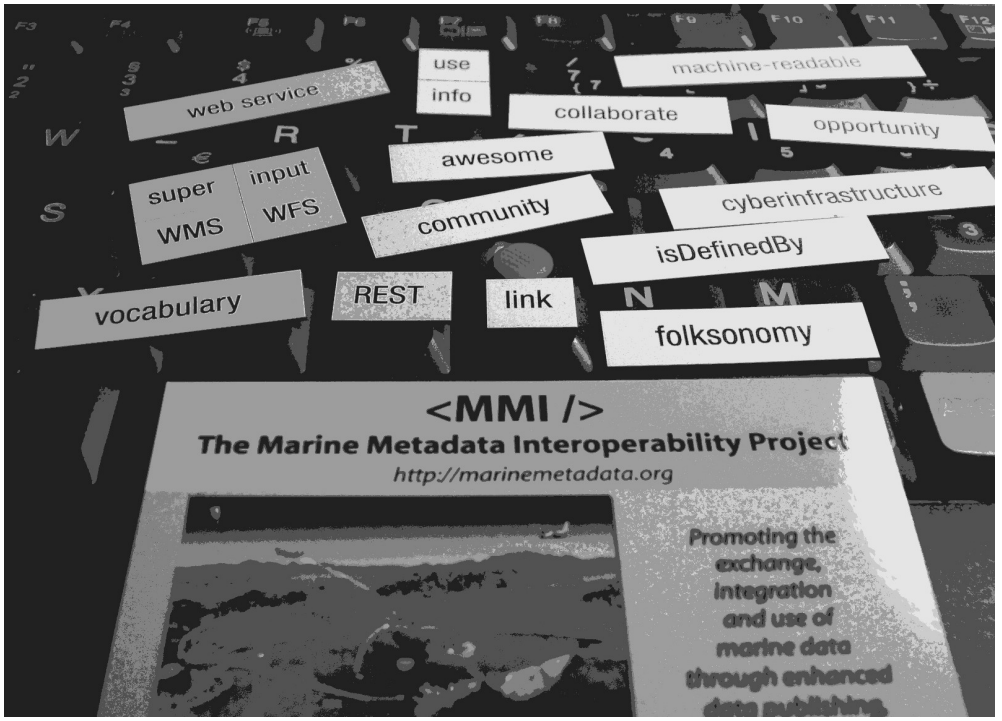
Ontology

In computer science and information science, an ontology is a **formal *representation*** of knowledge by a set of concepts within a domain and the **relationships** between those concepts.

It is used to **reason** about the properties of that domain, and may be used to **describe** the domain.

Ontology

Every Domain has to create their own Ontology.



Example:

The Marine Metadata Interoperability Project

http://marinemetadata.org

OGC

Ontology

A formal ontology is a controlled vocabulary expressed in an ontology representation language. This language has a **grammar** for using vocabulary terms to express something **meaningful** within a specified domain of interest.

The use of Ontologies is still limited!

Pragmatics

In Semiotics, **Syntax** and **Semantics** are complemented by **Pragmatics.** It defines:

>>> The **relation** between signs and their effects on those who use them.

>>> People, machines?

Informatics in general and geospatial data management e-spatially lack **Pragmatics**!

The Data Provider Perspective:

- All spatial data is described by the same smallest common denominator (Dublin Core, FGDC, ISO 19115 and so on) an specific incompatible profiles.
- Metadata creation and maintenance are typically artificial extra jobs for the data creators.
- Metadata is provided through complex interfaces (Catalog Services Web) and formats (ISO 19139 and specific profiles).
- Metadata quality is mostly limited on syntax.
- There is little to no interaction with users.

The Data Consumer Perspective

- Users do not understand the "language" of the providers
- Metadata descriptions are always incomplete if not related to one or better even – many ontologies.
- The service offering is too complicated, incompatible and unreliable.
- "Geo portals" do not satisfy user's need.
- Interaction between consumers and providers is minimal.
- Metadata is not linked well enough.

Nu Req'

- Extend offer
- Simplify Search.
- Allow browsing.
- ...user evaluation.
- Supersede categories by user's tags.
- Automize creation and maintenance of metadata
- Allow and enhance interaction between users and providers

Concepts of the Web (2.0)

- Resource-oriented architecture patterns (REST and ROA) allow simple creation, maintenance and search.
- Propagation of spatial data uses GeoRSS.
- User communities must grow their own specific ontologies.
- Users and providers need to talk.
- All data belongs in open buckets!

Filter on the way out!

David Weinberger (2008): Everything is Miscellaneous

OGC

REST – linking with sense

Four concepts:

- the resource
- the name (URL)
- the representation
- their relations(links)

Four properties:

- Addressability
- Statelessness
- Connectedness
- Well formed operations

The corresponding architecture pattern is the Resource Oriented Architecture (in short: ROA).

OGC

Now what?

Current catalog technology does not use the potential of metædata.

We need "buckets":

- **Open Access to Spatial Data and it's Metadata**

Now what?

The "Internet" is the lowest common layer of any SDI. We should use it as it was meant to be used:

- **REST paradigms and the ROA shall permeate standards (OGC, INSPIRE, ISO, etc.) and spatial offerings (OS UK, CGDI, Geoportal RLP etc.)**

Interested?

Questions?

Where to now?

~~Xxxxx~~ Denver 2011

~~2010~~ ~~XXXX~~

~~6~~ to ~~9~~ September ~~2010~~ 2011

~~XX~~ ~~XX~~ ~~XXXX~~

The international
Open Source Conference of the
Geospatial Domain

http://2011.foss4g.org

Powered by
Metaspatial