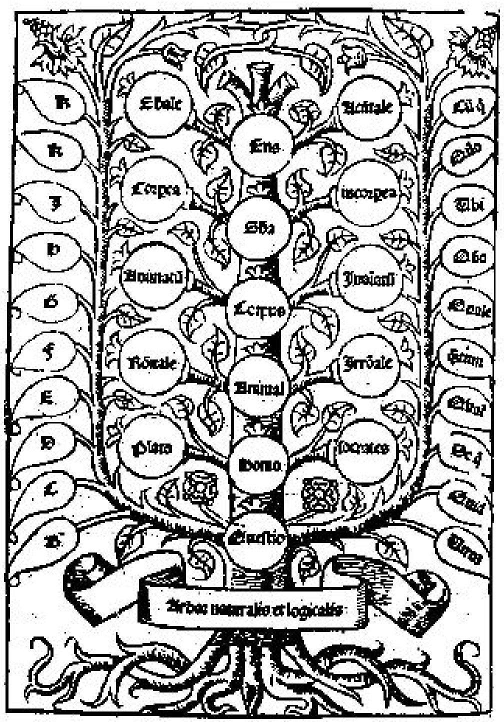


Visualization Ontologies:

Drawing pictures with meaning

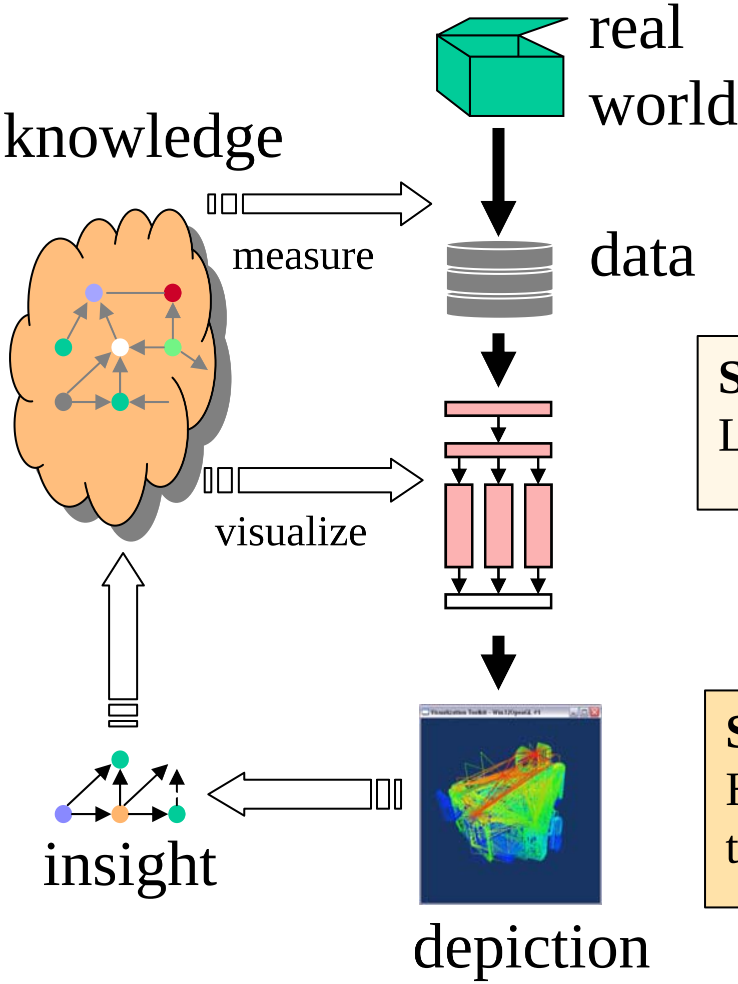
David Duke, Ken Brodlie

Visualization & Virtual Reality Group
School of Computing
University of Leeds

David Duce
Oxford Brookes University, Oxford

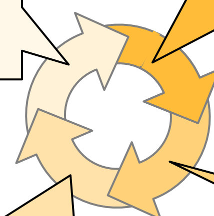
Ivan Herman
CWI, Amsterdam

- Motivation
  - How visualization works
  - e-Science requirements
  - Technical baseline
  - What is this ontology thing anyway?
- Building a visualization ontology
  - Initial sketch
  - Existing foundations
  - Web technologies
- Challenges



System to human:
Look at this!

Human to human:
How can we understand this data?



System to system:
Help to (filter | draw | ..) this data.

Human to system:
Visualize this!

- UK National e-Science Programme
  - large-scale collaborative, distributed, science
  - examples: Comb-e-Chem, AstroGRID, LGC
  - commercial grids

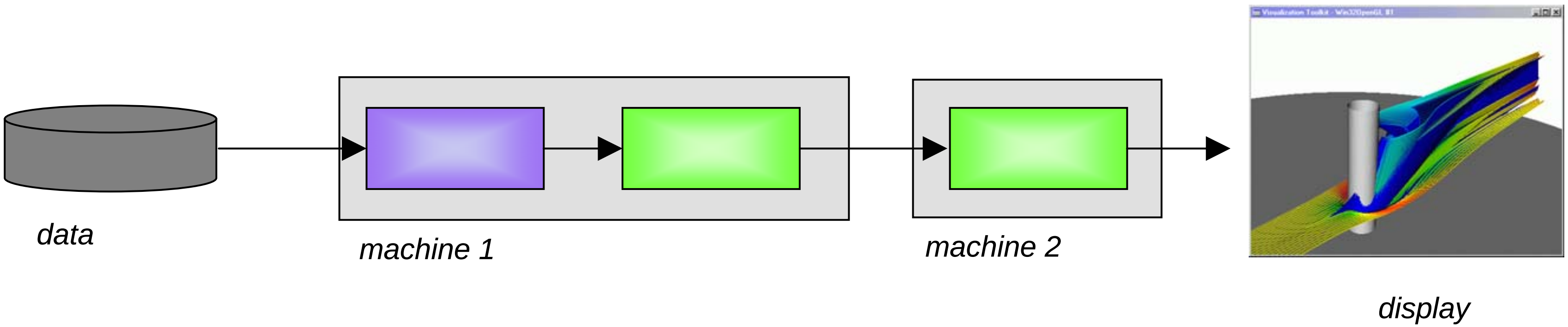
- Visualization is a significant component
  - how do you grid-enable vis? (see e.g. gVis, eVis, …)
  - how do domain specialists interface with vis?

- “Visualization for e-Science” Workshop, NeSC, Jan 03
  - “There is a need to establish an ontology for visualization. The development of this might be undertaken as part of the follow up workshop, but would need ongoing refinement before it could gain credibility.”

Roles for a *visualization* ontology

- Collaboration
  - common ground for for communication
  - what is an isosurface? … what is a cell?
- Composition
  - move towards visualization services
  - service discovery and (automated) assembly
  - map from conceptual task to physical resources
- Curation
  - experiments should be archived and must be reproducible
  - visualization often an important tool in analysis protocol
- Education
  - visualization as end user service
  - what is visual(ization) literacy?

Layers of abstraction



domain *"show the direction and temperature of fluid flow interrupted by the spike"*

conceptual *place streamlines into field, coloured by temperature*

logical *use streamline filter for lines; probe data along streamlines for temperature.*

physical
vtkStreamer >> vtkProbeFilter on machine1
-mpi-
vtkPolyDataMapper on machine2

What is (an) Ontology?

- Formally, it describes “what exists”
  - names concepts (defines a vocabulary of things)
  - names relationships (defines how things are linked)
  - relationships as concepts (reification)
  - concepts introduced
    - by assertion
    - by definition, wrt other concepts
- Not an entirely new idea …
  - taxonomies
  - database schemata
  - OOD
- … but given new life through the semantic web

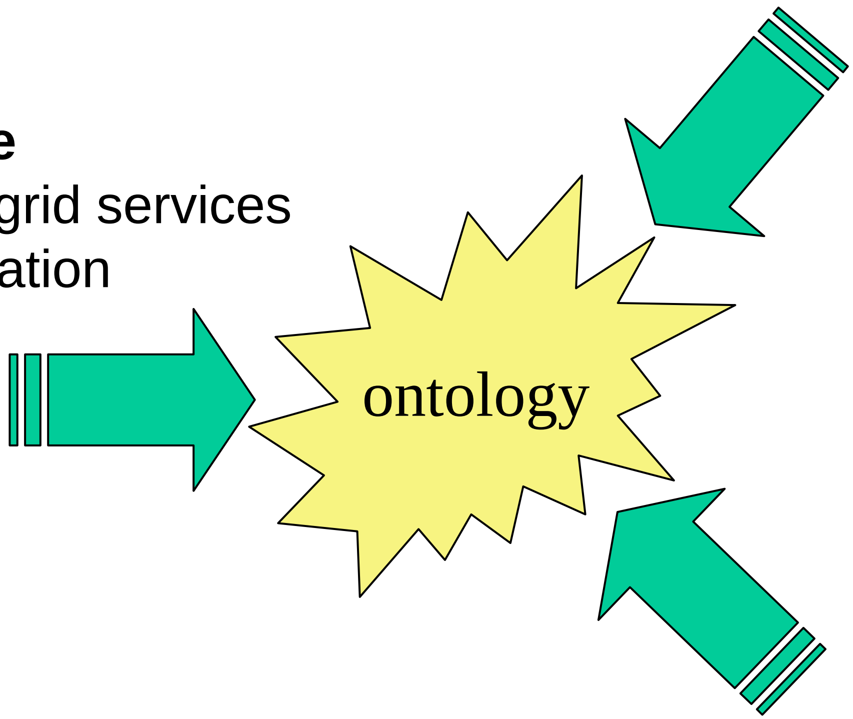
- The future of the Web
  - a universal medium for the exchange of data
  - personalized services (e.g., news)
  - integrating heterogenous Web databases
  - better search engines
  - adaptive presentations (e.g., for PDA-s, phones)

- The ***Semantic*** Web (IH)
  - machine processable metadata
  - unambiguous names for resources (URIs)
  - a common language for expressing metadata (RDF)
  - common vocabularies (Ontologies)

- *The "Semantic Web" is a metadata based infrastructure for reasoning on the Web*

Means
- OWL / RDF
- autonomics

Motive
- web/grid services
- education

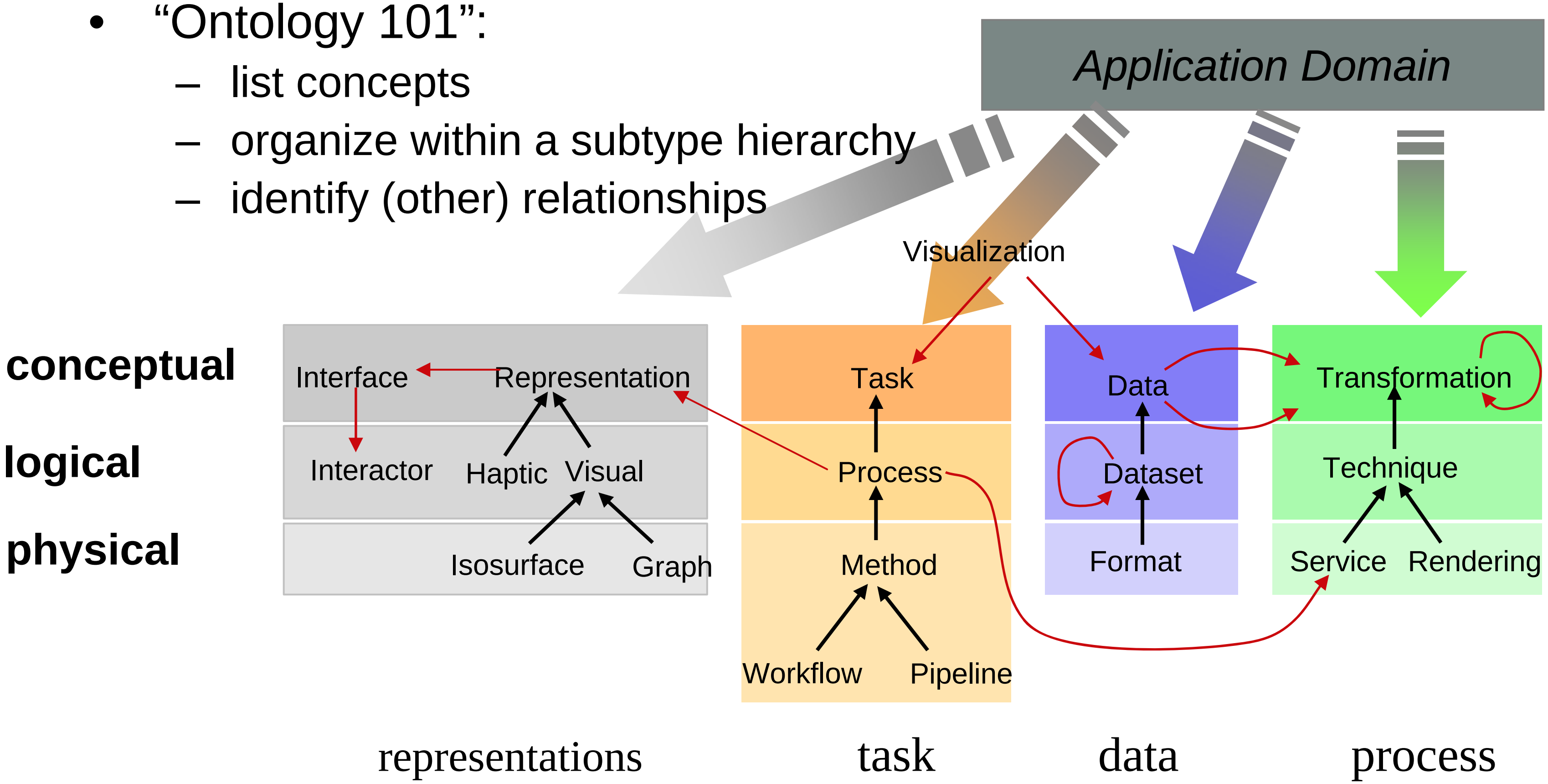


Opportunity
- e-Science.
- vis foundations

Building a Visualization Ontology

- Is an ontology for visualization feasible?
- What tools exist for building it?
  - ontology languages: RDF, RDFS, OWL
  - tools: Protégé, Triple20, …
- How might it be organized?
  - Levels of concern (from gVis)
    - conceptual: what is the visualization to be performed?
    - logical: what configuration of entities is required?
    - physical: how are entities bound to physical resources?
- Making progress:
  - top-down: use “Ontology 101” to sketch top-level organization
  - bottom-up: reconstruct from known taxonomies and APIs

- "Ontology 101":
  - list concepts
  - organize within a subtype hierarchy
  - identify (other) relationships



- Now relate categories to existing models and taxonomies

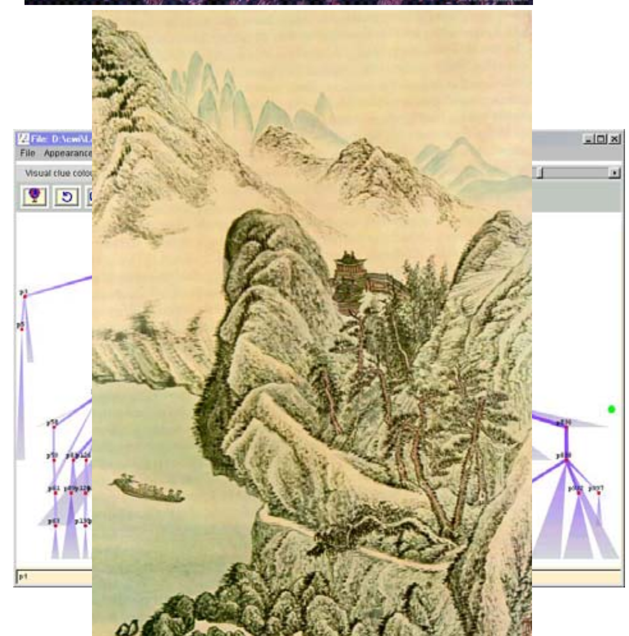
- Underlying field (Brodlie)
  - multi-variate, multidimensional function
  - $F = (f_1, f_2, \ldots f_m)(x_1, x_2, \ldots, x_n)$
  - For scivis, independent variables x_i are real
  - E- and O- notations ('92/'93)
- Terminology for real-valued dependent variables, $f_i(x)$
  - Scalar field: $f_i(x) \in R$
  - Vector field: $f_i(x) \in R^k$, $k \leq m$
  - Tensor field: $f_i(x) \in R^{k:k}$, $k \leq m$

- Structure of data sets
  - Butler and Pendley (fibre-bundle model, ‘89)
  - Bergeron et. al. (database model, ’94)
  - Treinish (fields, ’94)
  - Hibbard (lattice model, ‘95)
  - Card, Mackinlay and Schneiderman (readings in infovis, ‘99)

- Two groups of taxonomies
  - how is data organized?
    - dependent/independent variables,
    - type of mesh, etc.
  - relationship between the data and the underlying field
    - Brodlie (E-notation ‘93)
    - Tory and Möller (2004)

Representation

- Approaches build on semiotics
  - Bertin
  - Tufte
- Relationship between representation and data
  - Wehrend
  - Keller and Keller
  - Brodlie (O-notation)
  - Tory and Möller
- Work in computer depiction (NPR) and art
  - Willats, Durand
  - drawing system
  - denotation system
  - marks



- Bergeron's classification of goals

- Visual synthesis ("smart graphics")
  - Mackinlay
  - Casner
  - Feiner and Zhou

- Visualization tasks
  - Keller and Keller (Wehrend)
  - Tweedie
  - Chi and Riedl ("user operations")

- Work in other communities, e.g. CSCW

- Notion of visualization pipeline
  - conceptual: Haber/McNabb
  - logical: Upson et al (AVS paper)
  - physical: AVS, Data Explorer, VTK …
- Other models for visualization
  - spreadsheet
  - spray-rendering
  - blackboard systems
  - event-based composition
- More general approaches
  - workflow models (c.f. eVis)
  - OWL-S: composition model for web services

- Defining the ontology is only a start …
  - need to express it in machine-processable form
  - utilize semantic web technologies

- RDF (Resource Description Framework)
- Statements about *resources*
  - *Resources:* an element, a URI, a literal, …
  - *Properties:* directed relations between two resources

- *RDF Statements:*
  - "triples" of two resources bound by a property
  - usual terminology: (s,p,o) for subject, properties, object

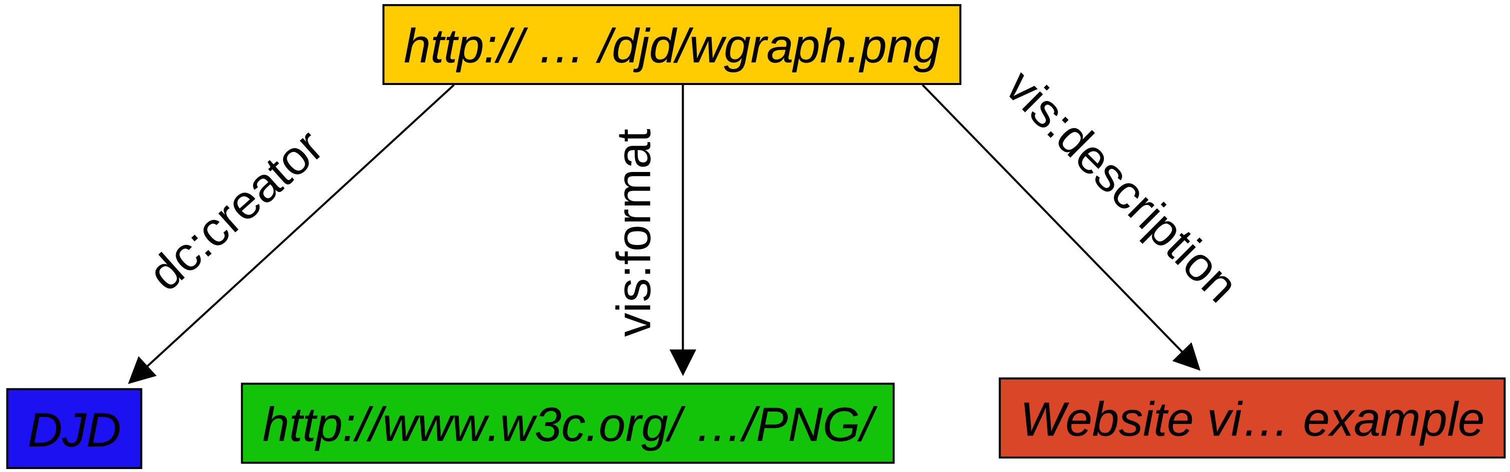
<rdf:Description rdf:about="http:// … /djd/wgraph.png">

<dc:creator>DJD</dc:creator>

<vis:format rdf:resource="http://www.w3c.org/.../PNG/">

<vis:description>Website visualization example</dc:description>

</rdf:Description>



- Adding metadata and using it from a program works…
- … provided the program *knows* what terms to use!

- However, what terms are known?
  - **creator, format, description?**
  - **owner, filetype, alternative?**
- RDF is the “machine code” of KR on SemWeb

- RDF Schema imposes some structure (ako Algol)
  - officially: “RDF Vocabulary Description Language”
  - sets out vocabulary for defining classes and relationships

- RDFS defines the terms of *resources* and *classes*:
  - everything in RDF is a “resource”
  - “classes” are also resources, but…
  - they are also a collection of possible resources (i.e., individuals)
- Relationships are defined among resources:
  - “typing”: an individual belongs to a specific class
  - “subclassing”: instance of one is also the instance of the other
    - similar to object-based programming…
    - … but the same resource can have several types
- “Type”, “subclass” are simple statements on resources
  - resources can be identified by URI-s
  - i.e., these statements can be described in RDF, too!

- RDFS is useful, but does not solve all the issues
  - Can a program *reason* about some terms? E.g.:
    - “if «A» is left of «B» and «B» is left of «C», is «A» left of «C»?”
    - programs should be able to *deduce* such statements
  - If somebody else defines a set of terms: are they the same?
    - obvious issue in an international context

- W3C’s Ontology Language (OWL) represents the next level
  - In RDFS, you can subclass existing classes…
  - … but, otherwise, that is all you can do

- In OWL, you can *construct* classes from existing ones:
  - enumerate its content
  - through intersection, union, complement
  - through property restrictions

Challenges

Caveats

- Monolithic ontology infeasible and undesirable
  - classification reflects pragmatic choices
  - fundamental differences between groups, e.g. ‘cell’
  - maintenance?
  - acceptance?
- Problem recognized within ontology community
  - “ontology islands” (Rousset)
  - seed process with with low(er) cost taxonomies and vocabularies
  - provide “bridging rules” between ontologies (Grau, Parsia & Sirin)
  - supporting technologies, e.g. SKOS (Miles, Rogers & Beckett)
- Managing the process
  - tension between openness and QA
  - open-source project approach?
  - oversight group – e.g. IEEE VGTC

- Engaging the community
  - BOF at IEEE Visualization 2005
  - Paper in IEEE CG&A
- Starting the work
  - Portal for SKOS vocabularies
  - RDF/OWL ontologies
  - (T)WIKI?
  - semanticweb.org