Deriving Relevant Information From 'Too Much Info'.

Authors:

Barbette Ivery - NAWCAD 4.5
Supporting PMA 290

Joe Schaff - AMEWAS Inc.
Supporting ACETEF

Semantics of Information = What Does it Mean?

- Raw data has no “real” meaning.
- Composition (creation of composite data) gives a meaning to the “composite data”.
  - **Traceable via decomposition**
  - **May mean more that its parts – synergy?**
- Framing and data linkages impart some aspects of meaning - i.e. the semantics to the data / metadata via its ***constraints***.

Example:

- **Two or more people have a conversation:**
  - ***several subjects covered***
  - ***numerous facts (= data items) mentioned.***
  - ***some are just opinions, possibly incorrect information.***
- **Now scale upwards - you are at a party:**
  - ***many conversations between two or more people.***

What's a first step?

- ***You do "conversation surfing":***
  - listen to snippets from one conversation, then move to next.
  - observe tone of voice, body language - emotional aspects.
  - gather information, <u>not</u> specific data.

How do we capture the implicit aspects of the conversations?

- ***We assess what we hear:***
  - party perspective – is there a theme (if any), general tone?
  - is there something external to this party that collectively affects the feeling - e.g. the economy?
  - do they "project" a common belief or feeling?
  - how do you form a composite picture of all that transpired in the conversations?
  - What is truly relevant information to take away from the experience?
- Each conversation had a unique flow to it, you were able to capture snippets of it.
  - from snippets you developed your perspective.
  - you want as unbiased as possible within scope.

What if we had thousands of conversations?

- **We would have thousands of functions. That's a <u>problem</u>.**

General Questions on Data –

bounding, framing, coupling, cohesion

- How does the data fit?
- How does new data “map” to the correct reference frame?
- Does the meaning of the data change when it becomes “framed”?
- Do current methods fully address the semantics of information w.r.t. data?
- How is data constrained?

Information Bounding – Cohesion and Coupling

- Best ***cohesion*** if data is grouped by task, sequence, or communication.
- Worst = logical (function-specific –e.g. radio knob and r.f. packet data) or most-commonly-used. Many data architectures leverage off of logical, etc.
- ***Coupling*** is the dependency of system / sub-element on others. ***Least is best*** .
  - Why? Check out compsci. refs. on program modules, or distributed / ‘cloud’ computing.
    - Also: comm channels, w/ “common” data shared.

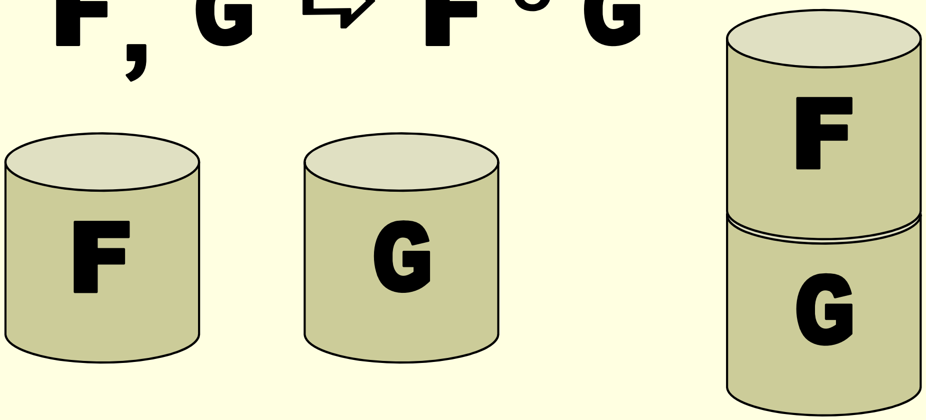
Optimize Coupling and Cohesion

Framing the Data

Creating the frame:

- Define reference frame.
- Relevance within frame.
- Linkages to other reference frames.
- Composite data Boundaries

F, G ⇨ F ° G



Caveats to address:

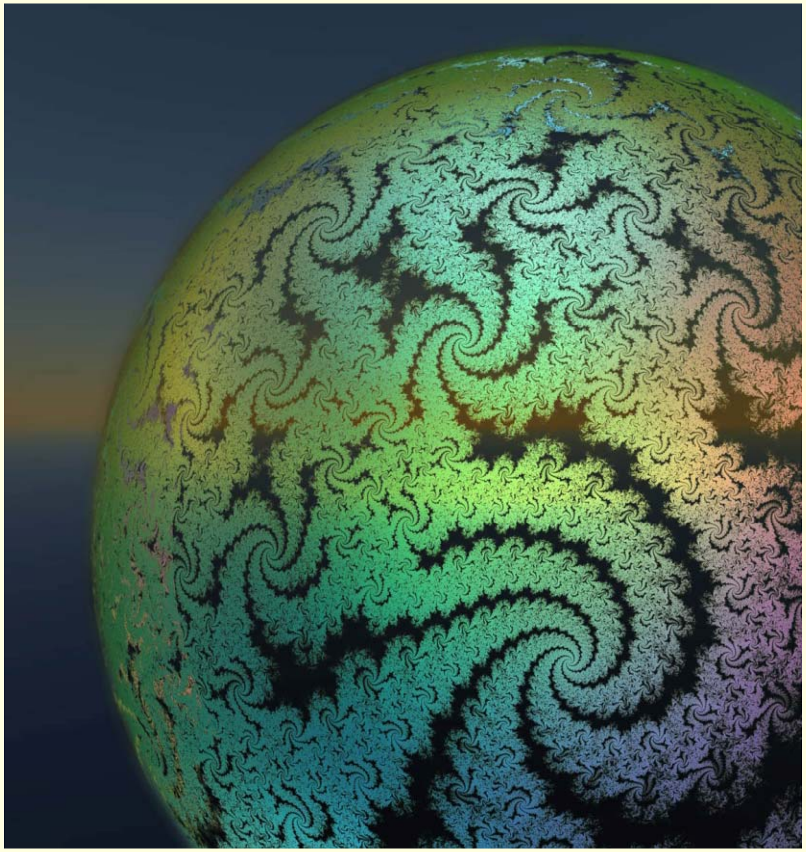
- Composite data may fall outside reference frame - e.g. F and G are both within frame, but F*G may be outside of it.
- Frame boundaries may change.
- Implicit linkages may not be visible.

Current Methods & Way Forward

- Methods include direct interpretation of triples (subject –predicate-object), latent semantics (implied meaning).
- Latent semantics / other methods are prone to error (chirality, causal-chains, etc.).
- Need mathematical formalism w/ consistent mappings.
- Need to capture implicit & explicit relationships between data, both spatial and temporal for cause-effect chains.
- Use clustering & topological mappings to show relationships.
- Use a vector-like construct to capture ‘causal’ linkages between information sets or groups.

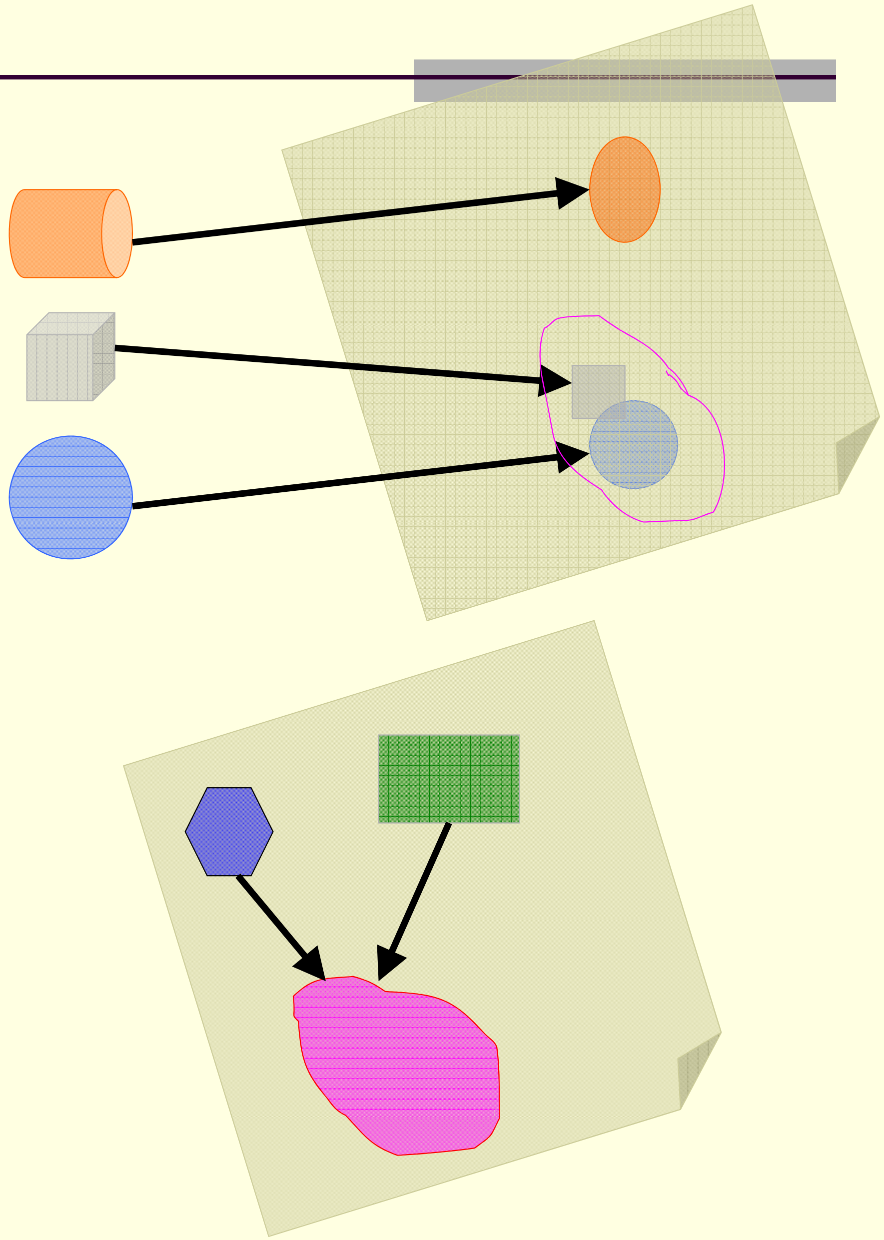
How does it work?

- Each unique “metadata” cluster has information that can be represented by parameters – which we call a function.
- Many of these distributed clusters can be grouped into a “space” of high dimension, typically referred to as ‘n-space’ or Hilbert space (n = number of dimensions).
- The n-space has the sum total of our needed information, but is too complex to work with.
- **Now what???**



How does it work (part 2)?

- We use mathematical topology to our advantage – and *project* the selected relevant information onto a plane.
- This data forms unique shapes that define relationships (e.g. Category Theory).
- We can also link two or more of these shapes to show cause-effects or other time-domain relationships.

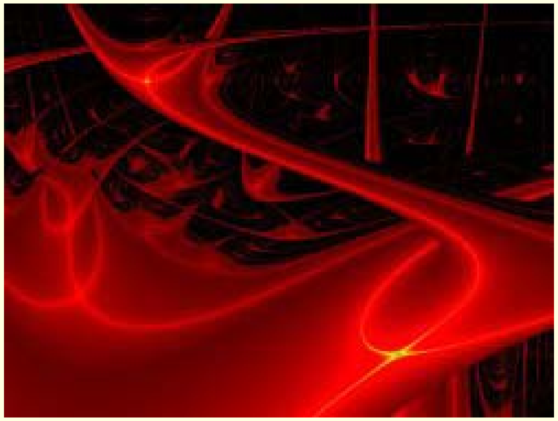


How does it work (part 3)?

From Abstract Mathematics to Language:

- The shapes and their respective ordering relationships form a grammar that describes events, entities, and effects via information flows.
- Now the appropriately framed data has an emergent meaning – i.e. semantics unique to the info-space *perspective*.

This “emergent” semantic perspective can be represented by a mathematical equation which could depict a seemingly abstract shape, below:



Conclusion

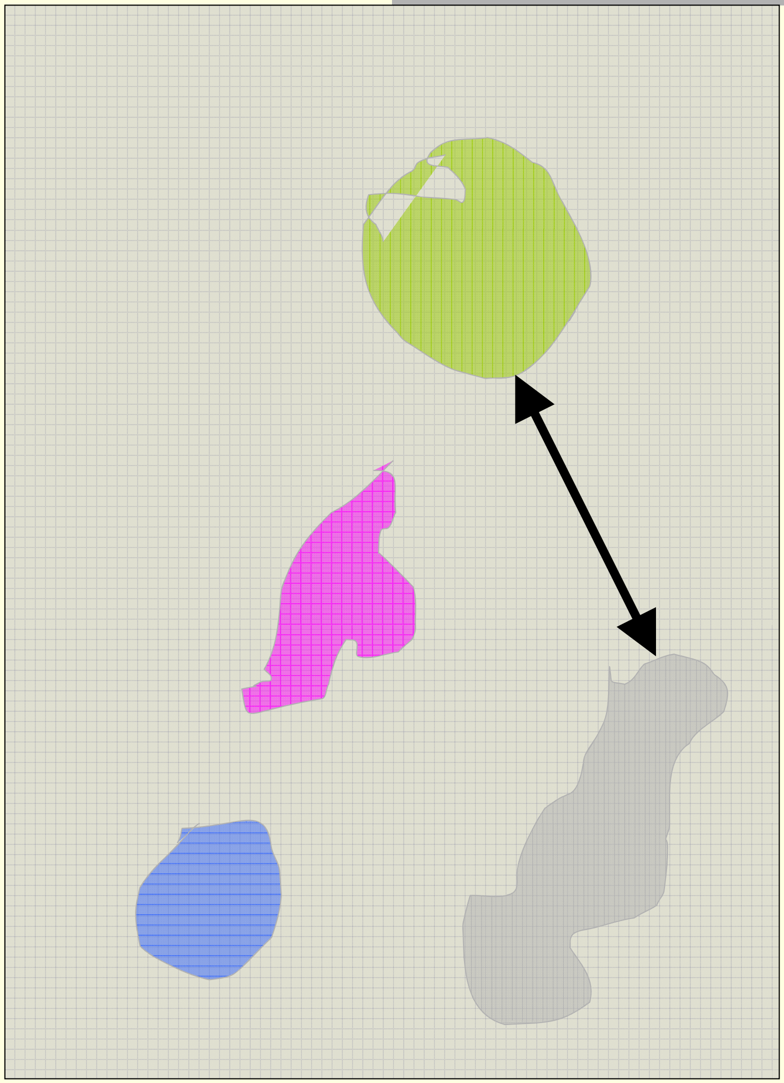
- ***Many applications regarding information validity and relationships. We can now address problems such as:***
  - ***Mitigation of friendly fire***
  - ***Heterogeneous large scale information integration – Army's F.C.S.***
  - ***Readiness metric and overall "Big Picture" (COP) of {capability, availability, dynamic resources effects}***
- Explicit and implicit meaning of data can be defined by selectively bounding clusters to form relevant information.
- Mathematical formalisms give credence to the results, and leverage off of topology.
- Newly 'discoverable' information becomes available due to its emergent properties.
- Both time-domain and causal (cause-effect) information is available using this method of information fusion.

Backup slide: “Comm. Channel”

- A generic communication channel spans the difference in time and in space separating senders from receivers. Comm. channel characteristics are:
  - the physical properties of its medium imposes a constraint on the capacity for communication
  - a specific capacity to store, retain, and transmit certain kinds of signals
  - a sensitivity to non-systematic distortions and decay (noise, etc.)
- A more generalized definition of the comm. channel can be applied to behaviors:
  - Behavior of a system within an environment has effects on other systems in the environment - the environment imposes physical constraints on ‘capacity’ of the system.
  - The capacity to affect or be influenced by effects on a shared behavior channel.

Backup2: Category Theory

- What is it?
  - a means of formally capturing mathematical structures by defining the structure-preserving functions that connect them.
  - Focus on the structure-preserving mappings between groups of objects (called “functors”).
- How does it fit?
  - shape-grammar projections have vectors connecting shapes, showing relational mappings.



Backup3: Barnsley Fern

An iterative Function System (IFS), is generated using the four equations:

$$f(x,y) = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.16 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$f(x,y) = \begin{bmatrix} 0.85 & 0.04 \\ -0.04 & 0.85 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 1.60 \end{bmatrix}$$

$$f(x,y) = \begin{bmatrix} 0.20 & -0.26 \\ 0.23 & 0.22 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 1.60 \end{bmatrix}$$

$$f(x,y) = \begin{bmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 0.44 \end{bmatrix}$$