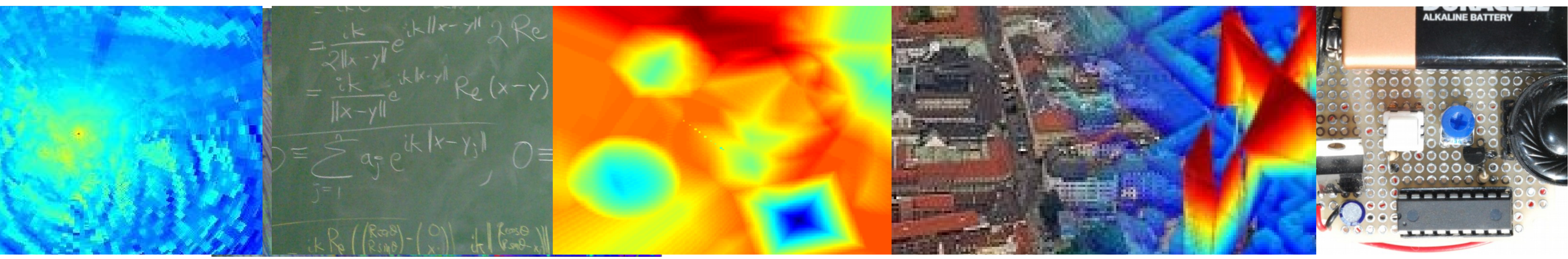
Sheaf methods for inference



Michael Robinson

Acknowledgements

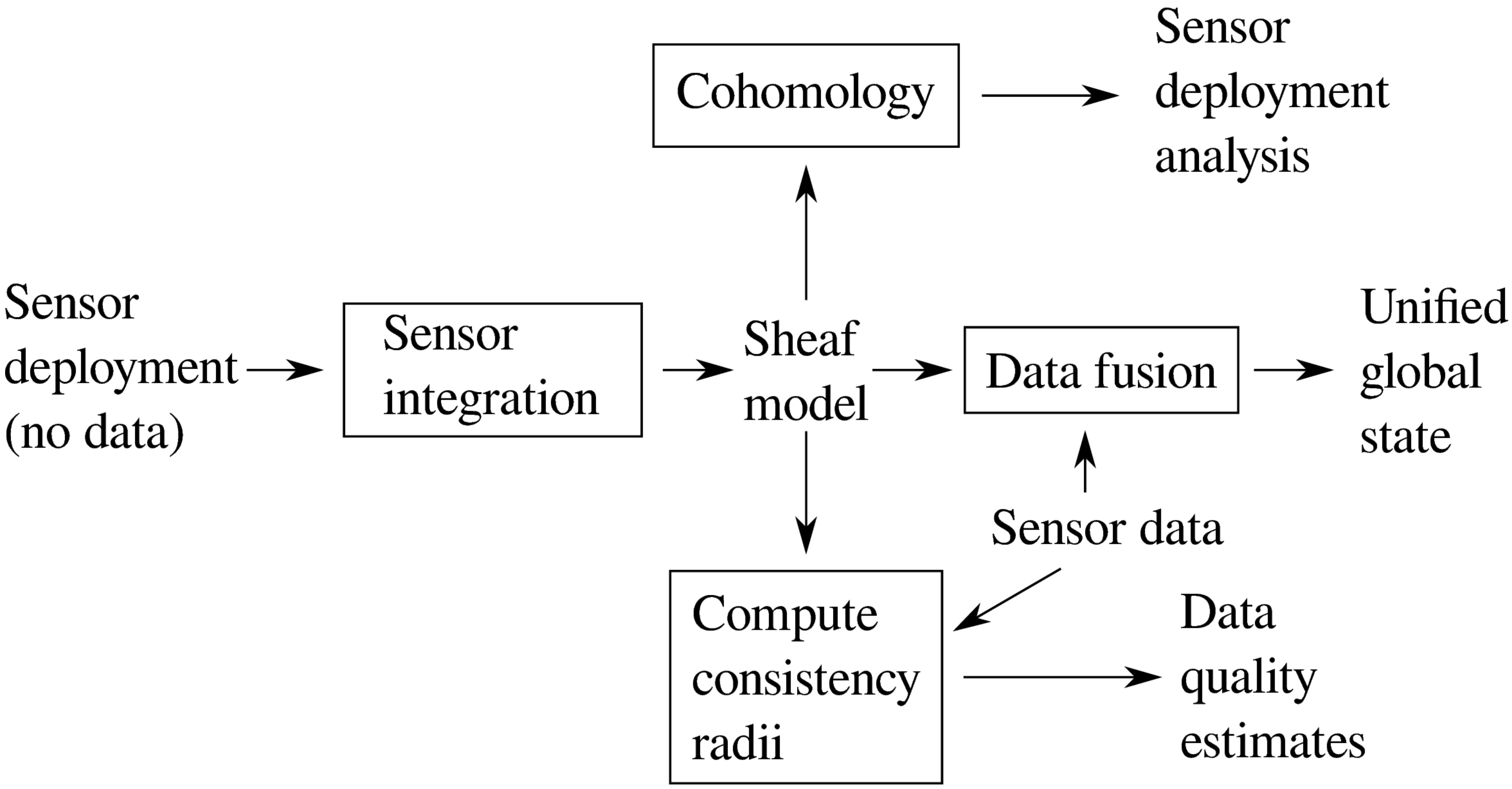
- Current students:
  - Eyerusalem Abebe
  - Phil Coyle
  - Brian DiZio
  - Jen Dumiak
  - Sean Fennell
  - Jackson Williams
- Funding: AFOSR, DARPA, ONR, OSD, AFRL, DTRA, PNNL
- SRC
  - Chris Capraro
  - Janelle Henrich
  - DJ Isereau
  - Sean Sweeney
- PNNL:
  - Cliff Joslyn
  - Katy Nowak
  - Brenda Praggastis
  - Emilie Purvine

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Three main ideas in this talk

1. Model multi-way relations with **topological spaces**
2. Model consistency between observations with a **sheaf**
3. Measure observation quality by projecting them onto **sections** of that sheaf

Topological processing workflow



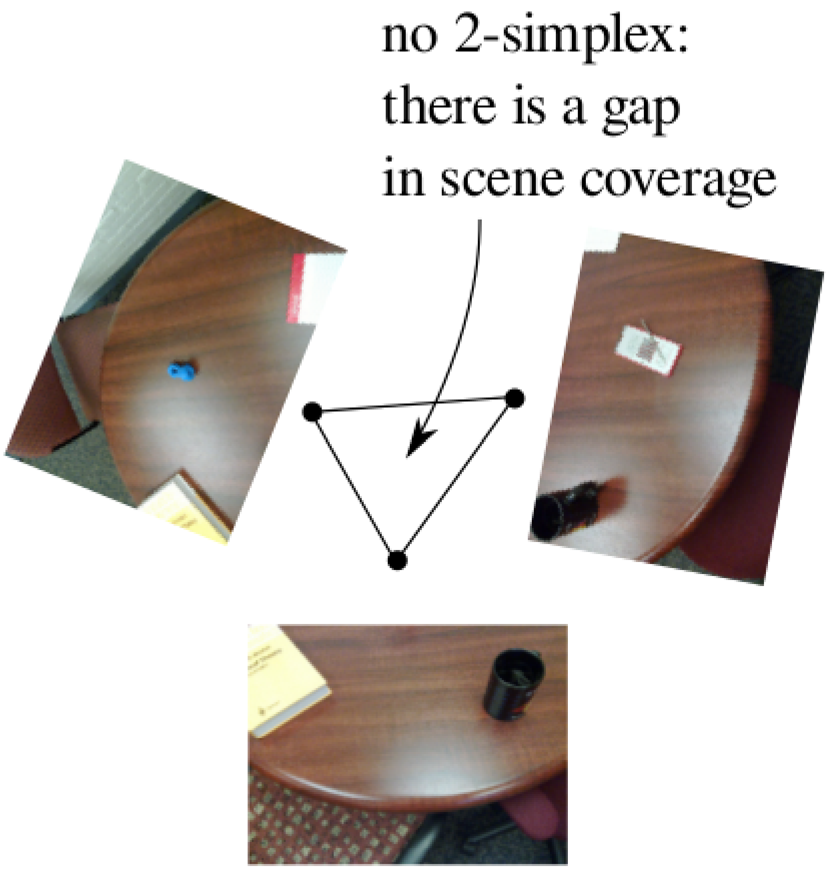
What is a sheaf?

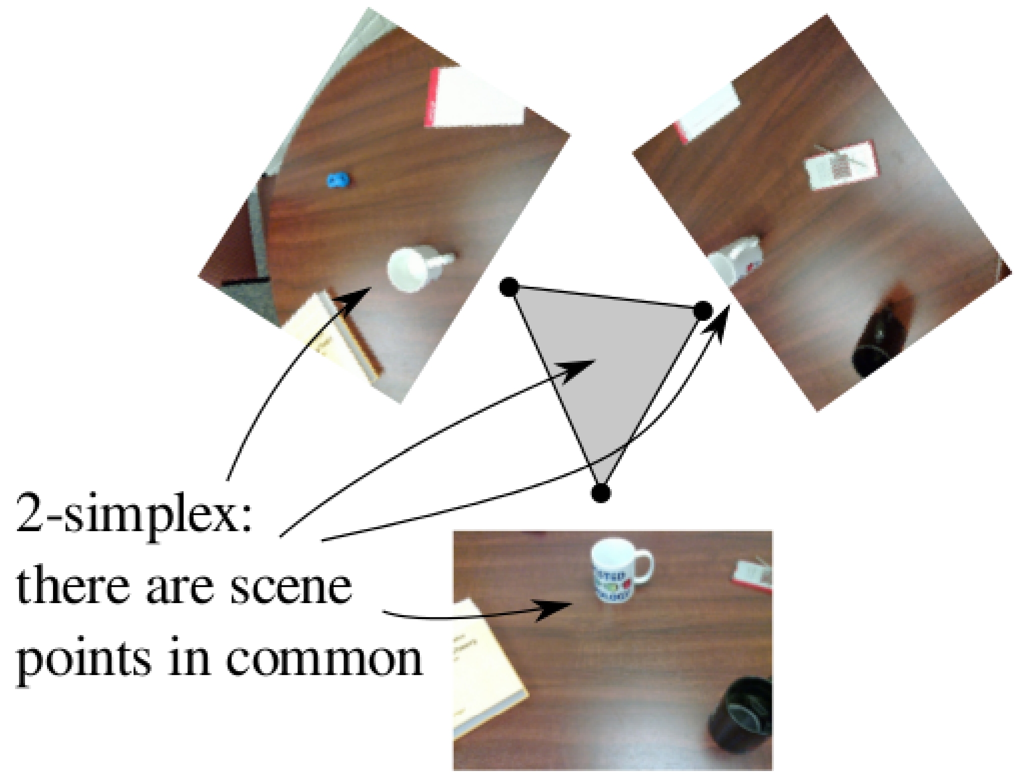
A *sheaf* of ________________ on a _________________

(data type) (topological space)

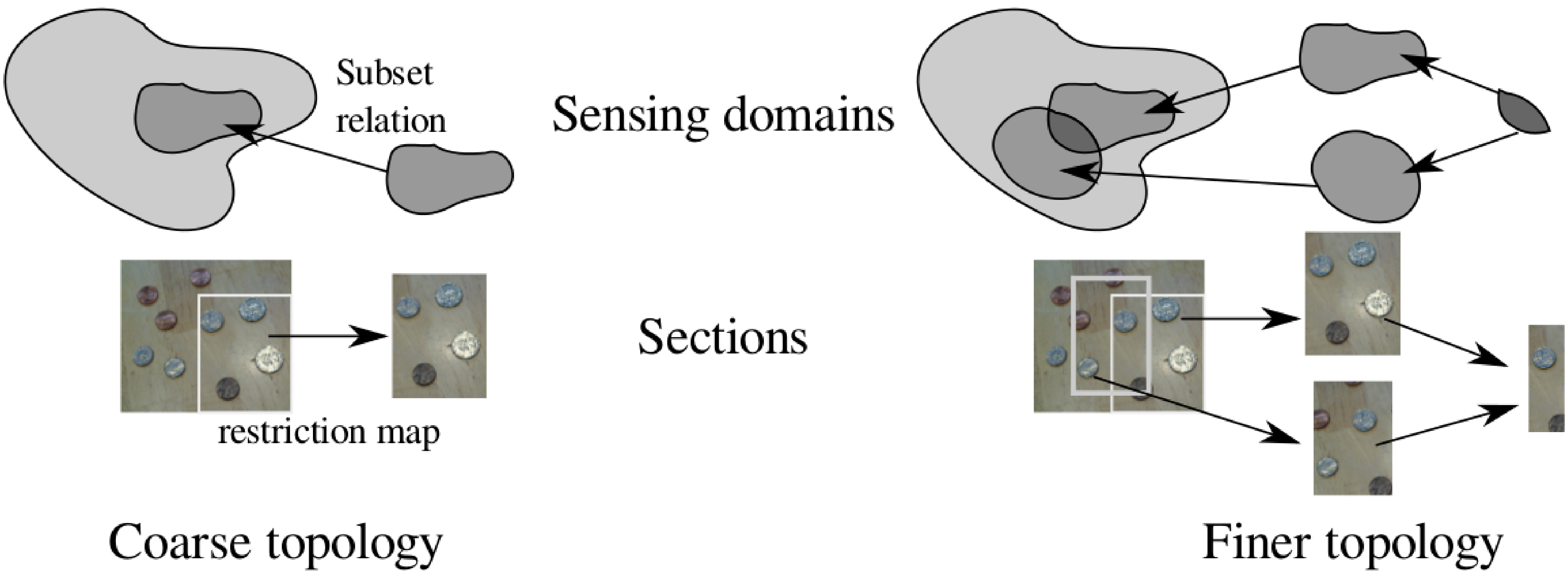
Michael Robinson

Overlap constructs topology

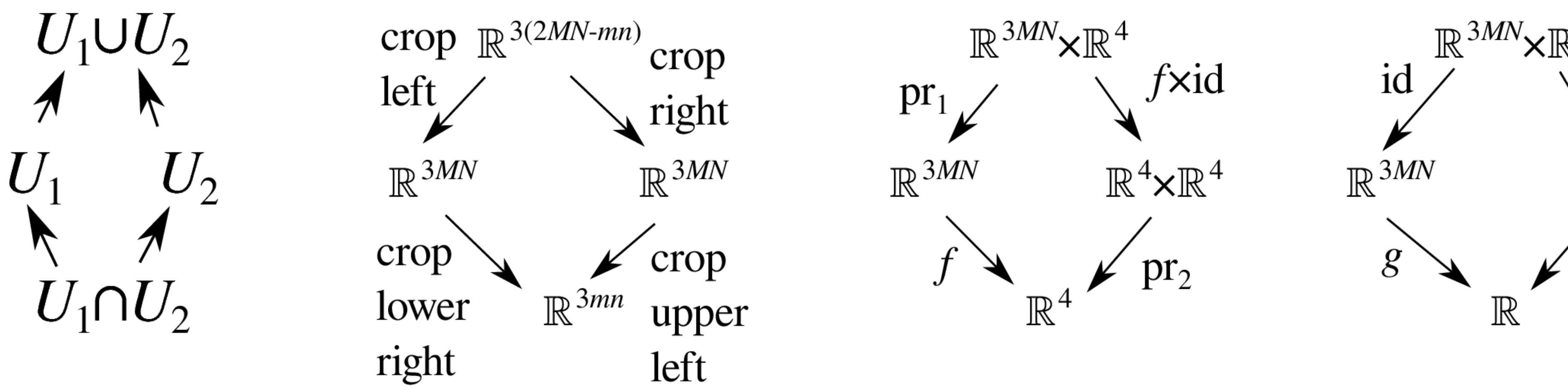


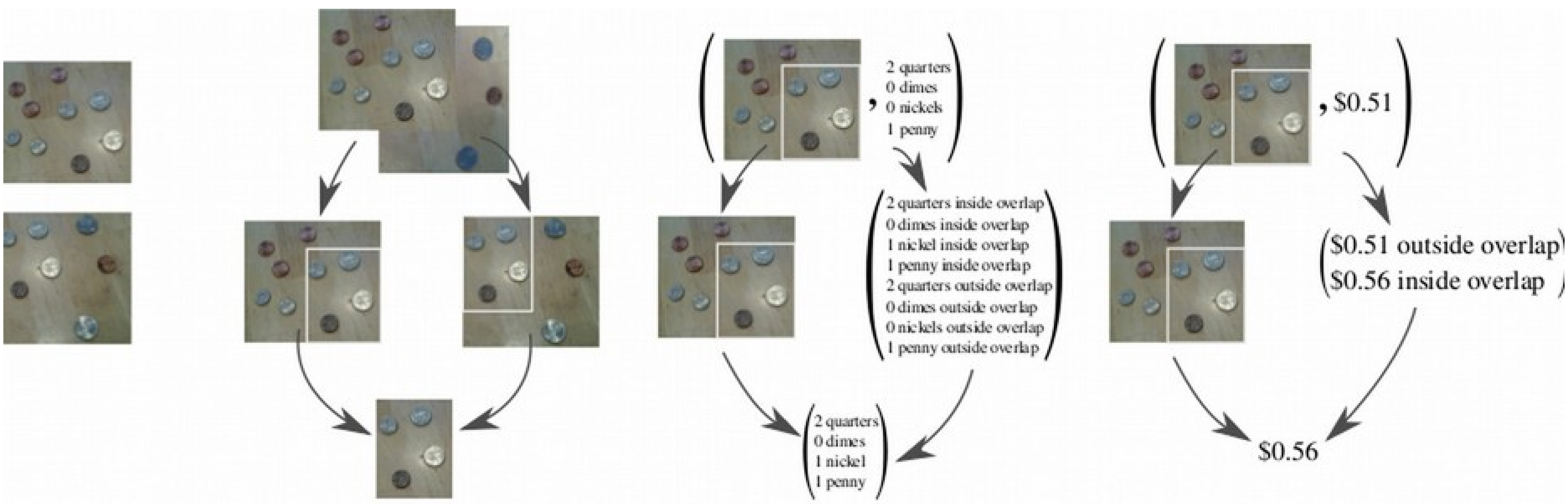


Changing overlaps changes the topology



Sheaves are about consistency

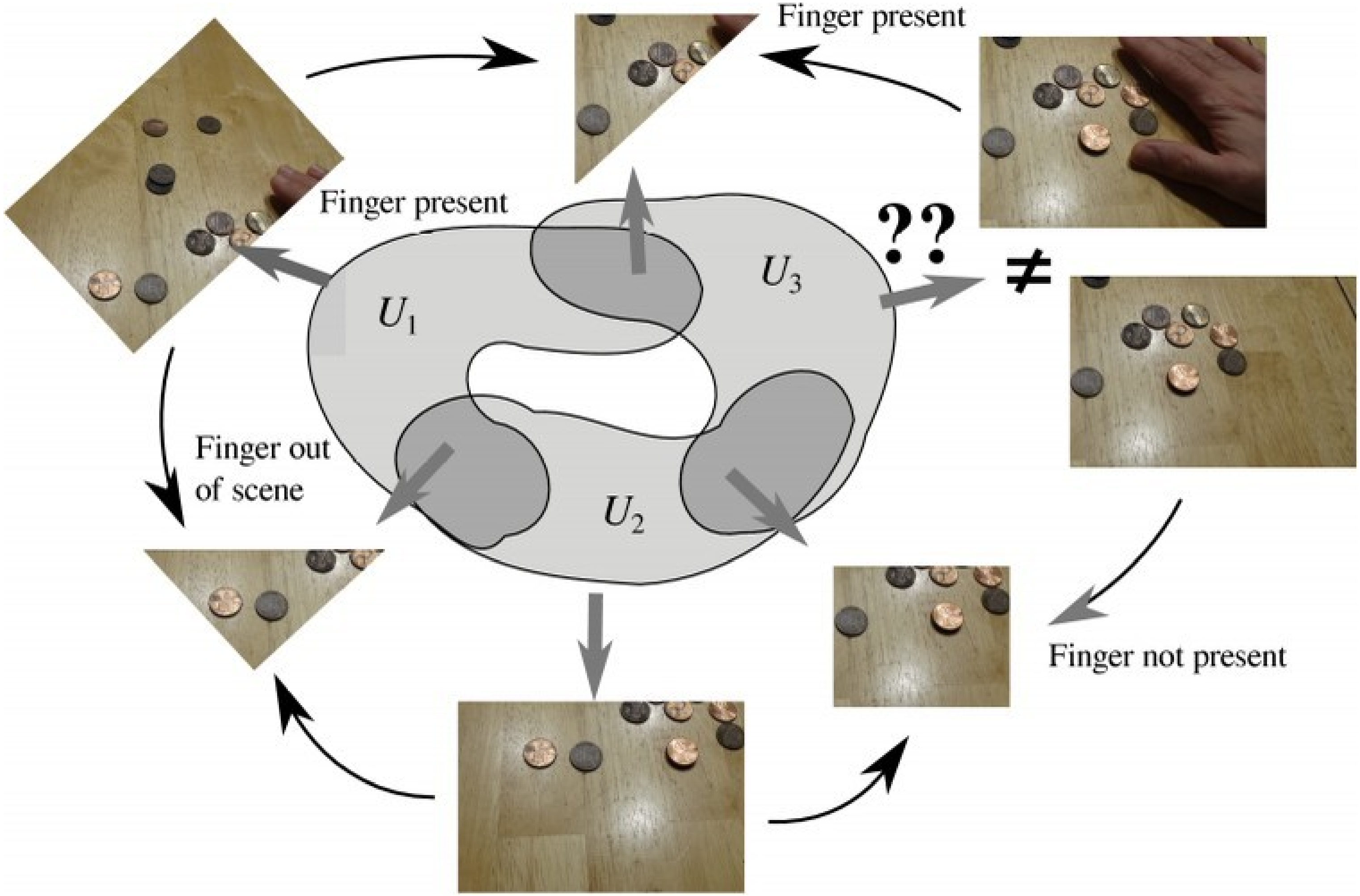




Non-numeric data types of varying complexity can certainly be supported! (Emilie Purvine's talk!)

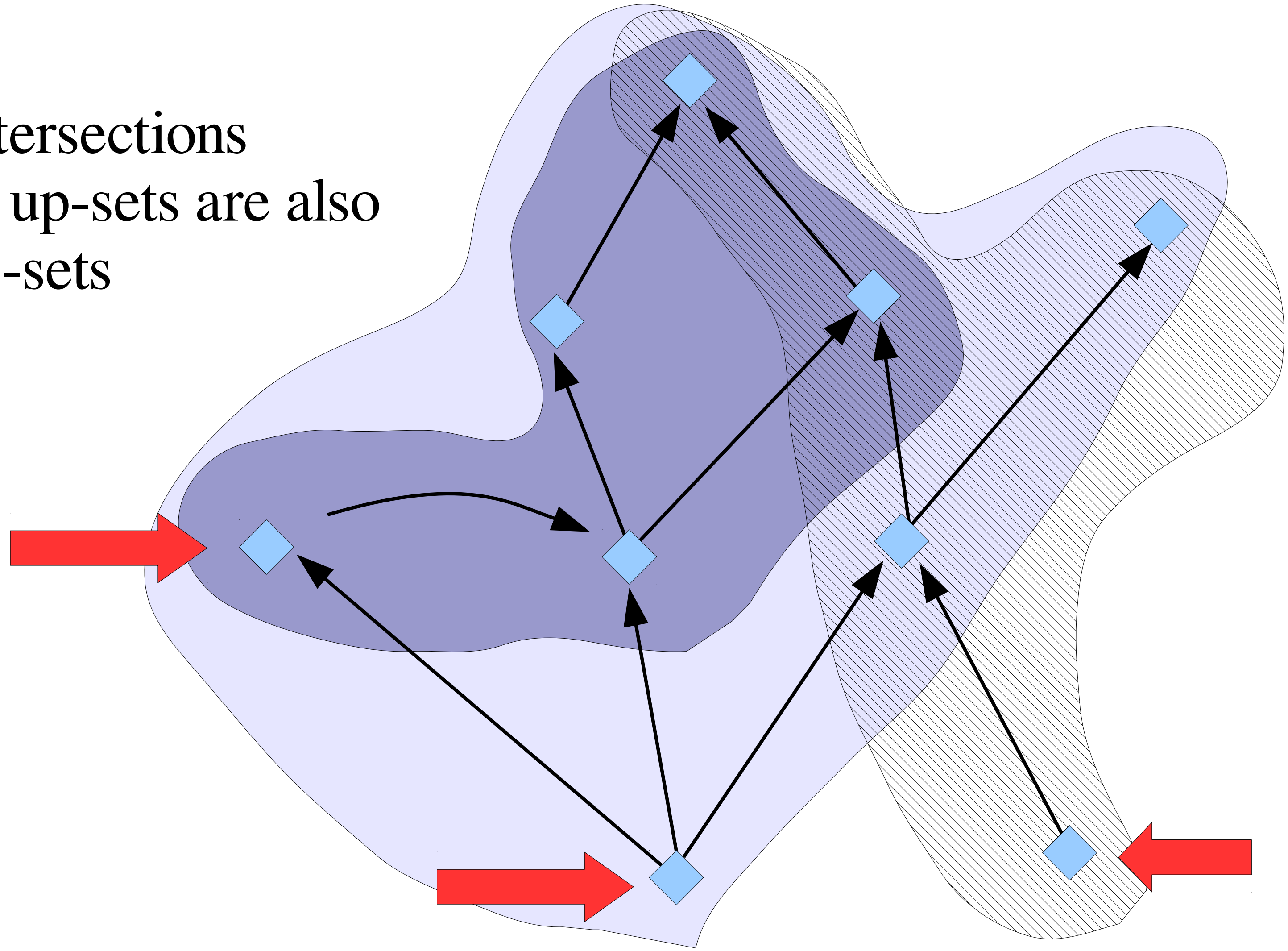
Michael Robinson

Data self-consistency and quality



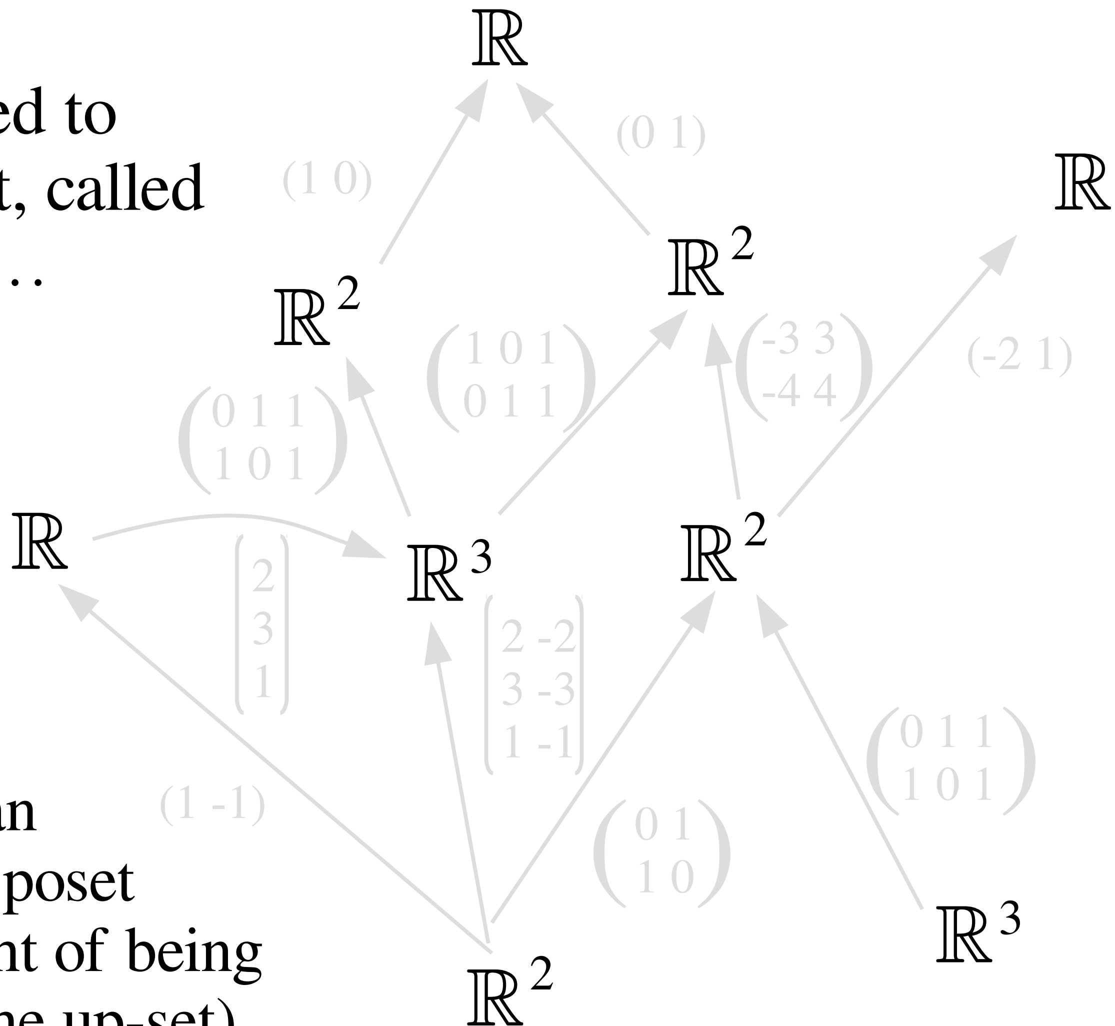
Topologizing a partial order

Intersections of up-sets are also up-sets



A *sheaf* on a poset is...

A set assigned to each element, called a *stalk*, and …

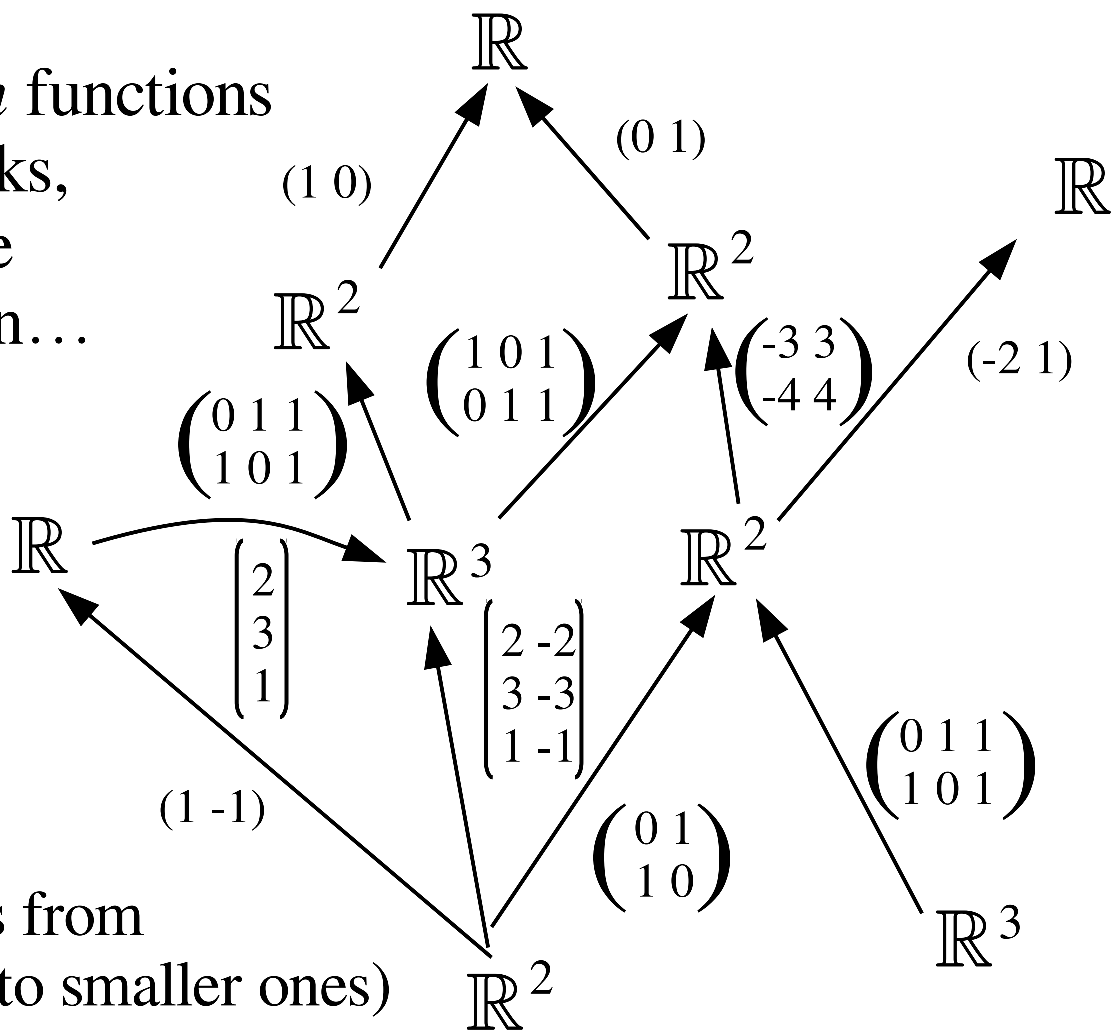


(The stalk on an element in the poset is better thought of being associated to the up-set)

This is a *sheaf* **of** vector spaces **on** a partial order

A *sheaf* on a poset is...

… *restriction* functions between stalks, following the order relation…

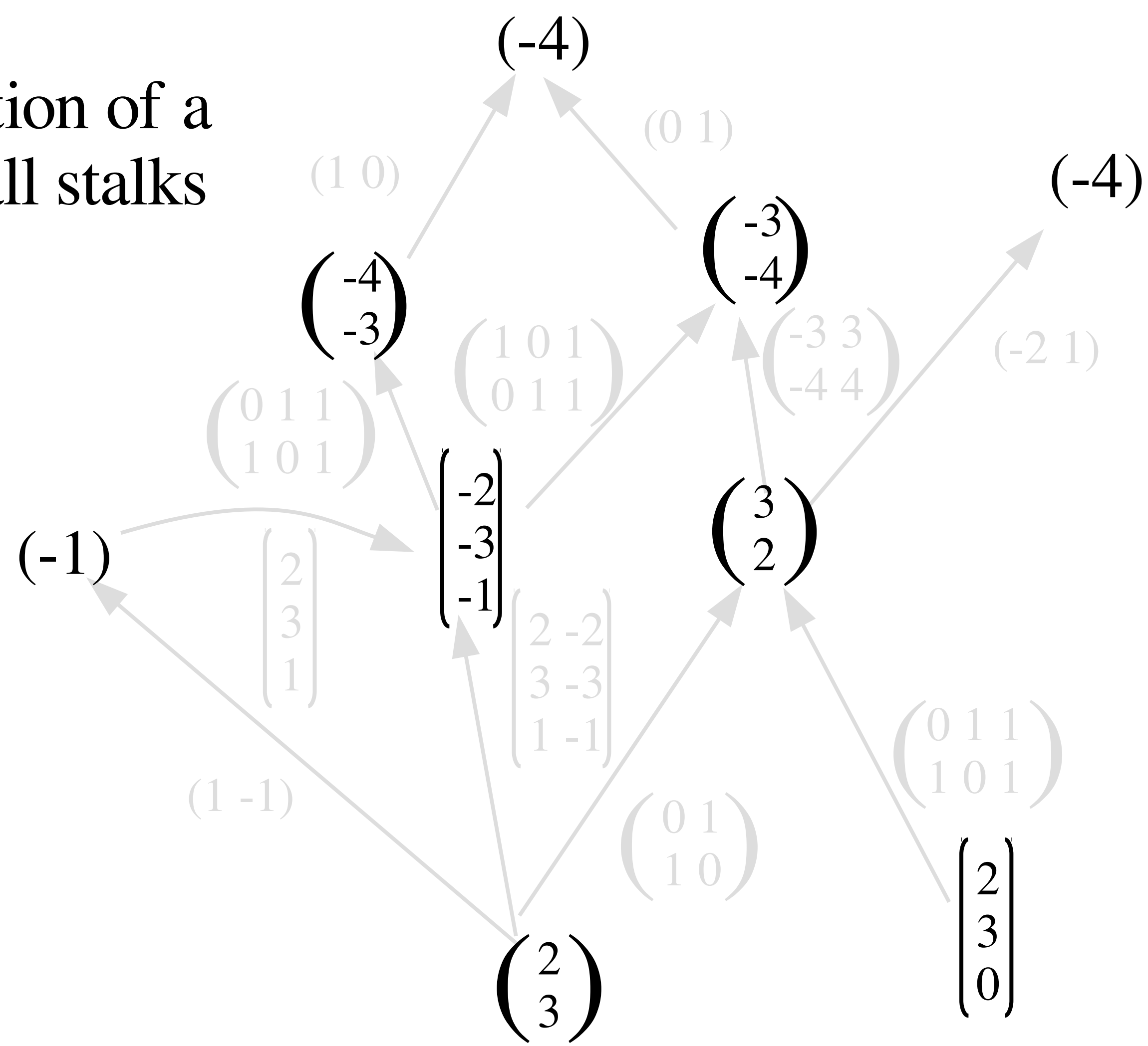


("Restriction" because it goes from bigger up-sets to smaller ones)

This is a *sheaf* **of** vector spaces **on** a partial order

An *assignment* is...

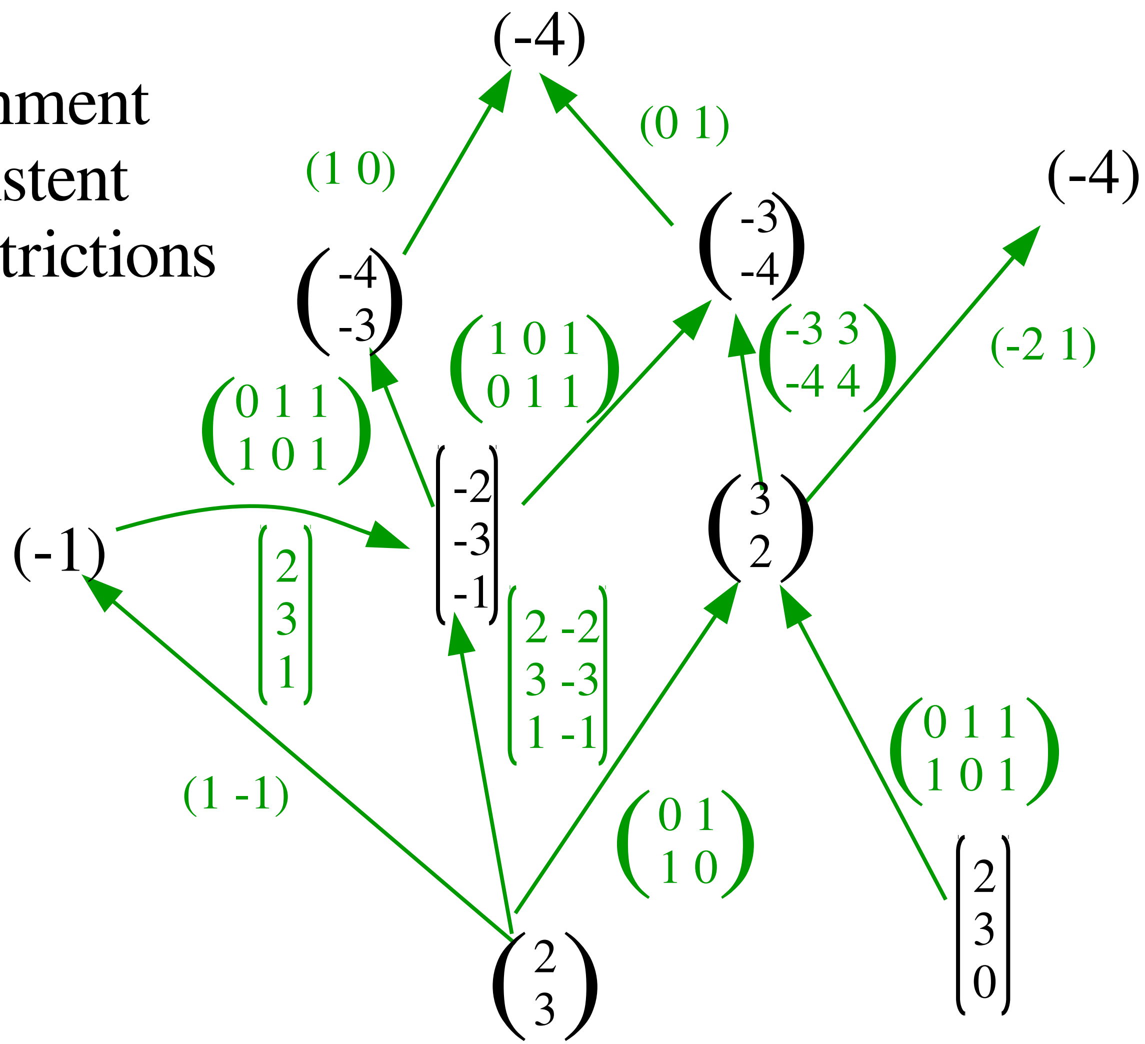
… the selection of a value from all stalks



The term *serration* is more common, but perhaps more opaque.

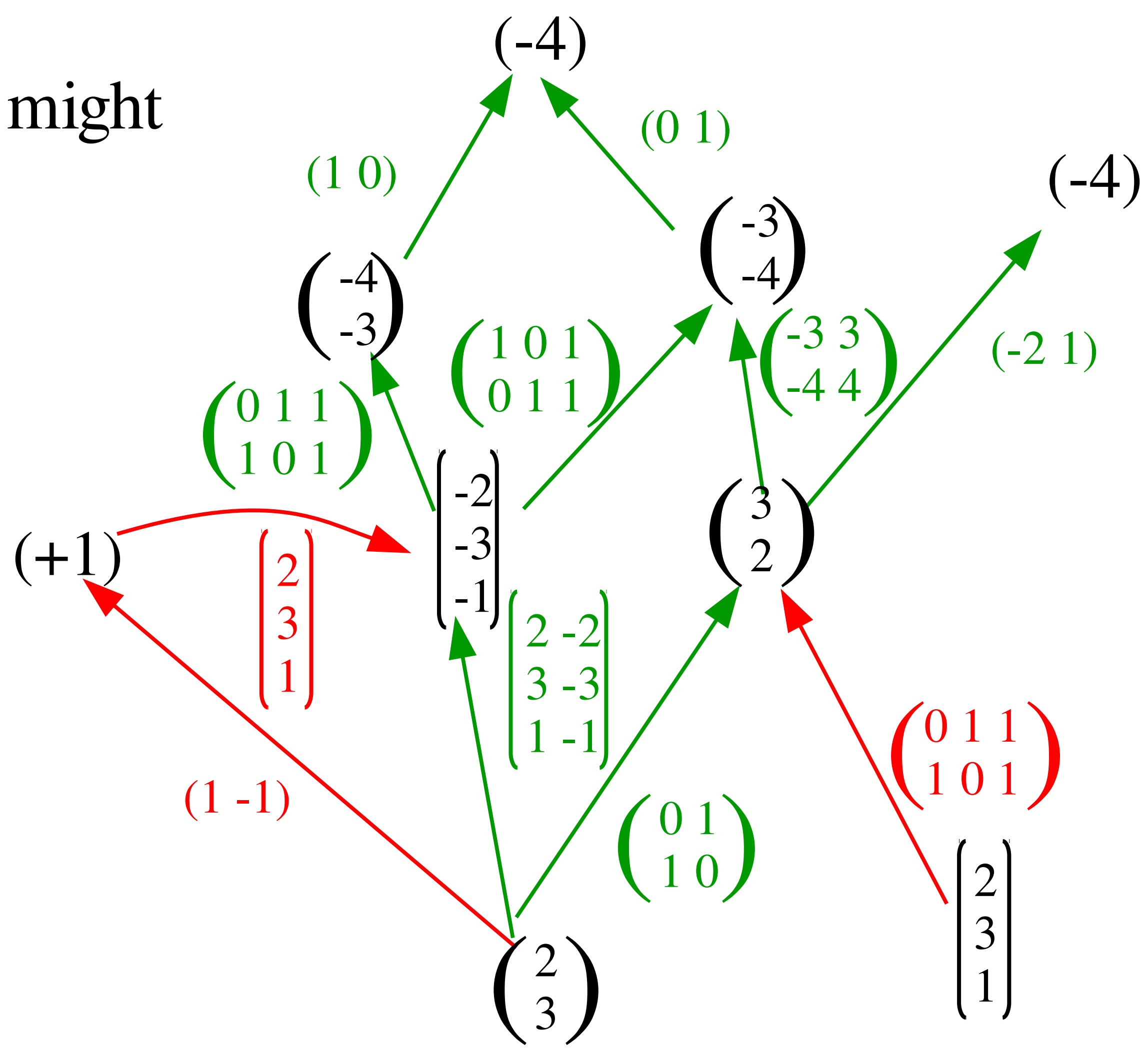
A *global section* is...

… an assignment that is consistent with the restrictions



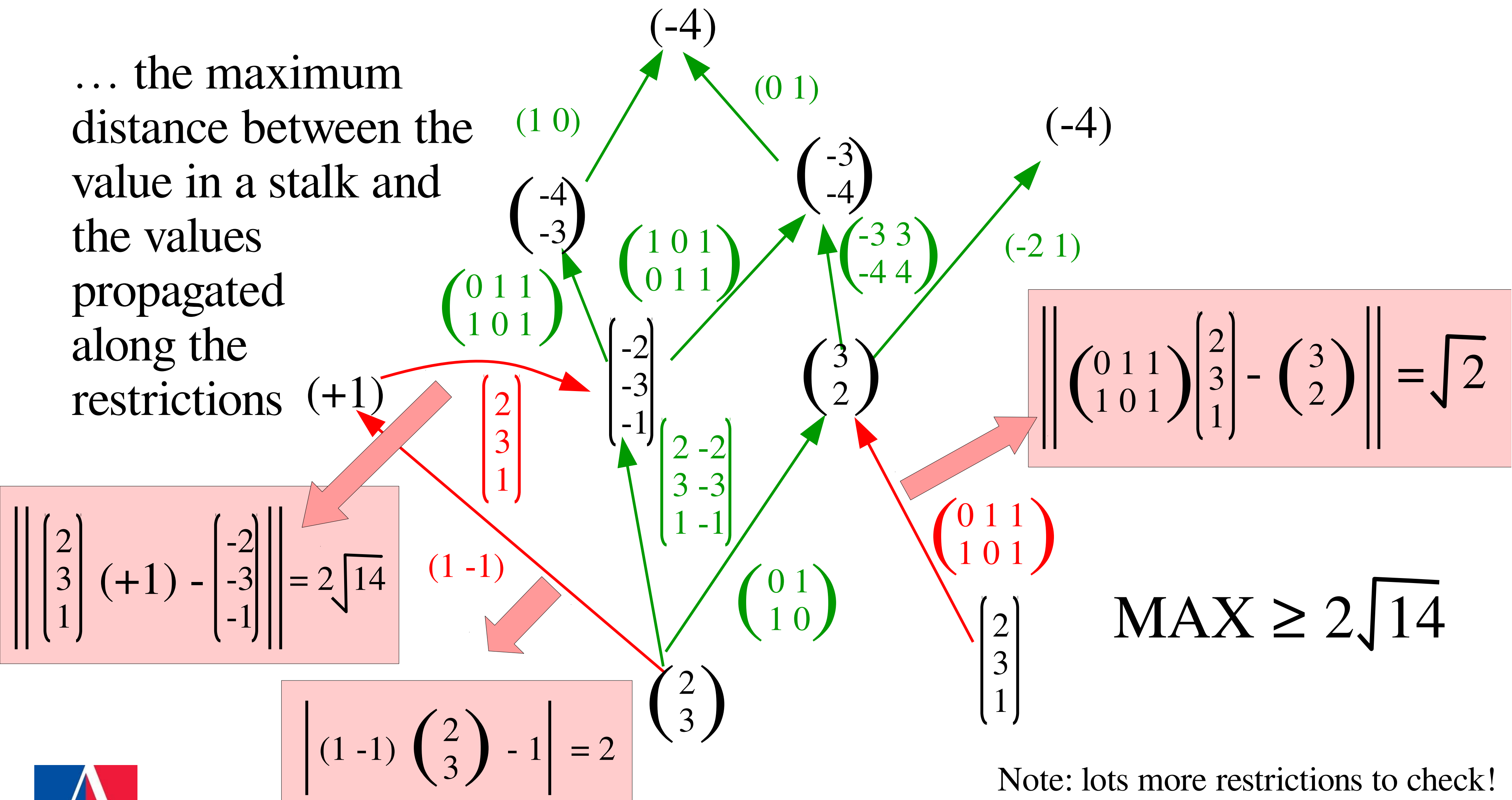
Some assignments aren't consistent

… but they might be partially consistent



Consistency radius is…

… the maximum distance between the value in a stalk and the values propagated along the restrictions

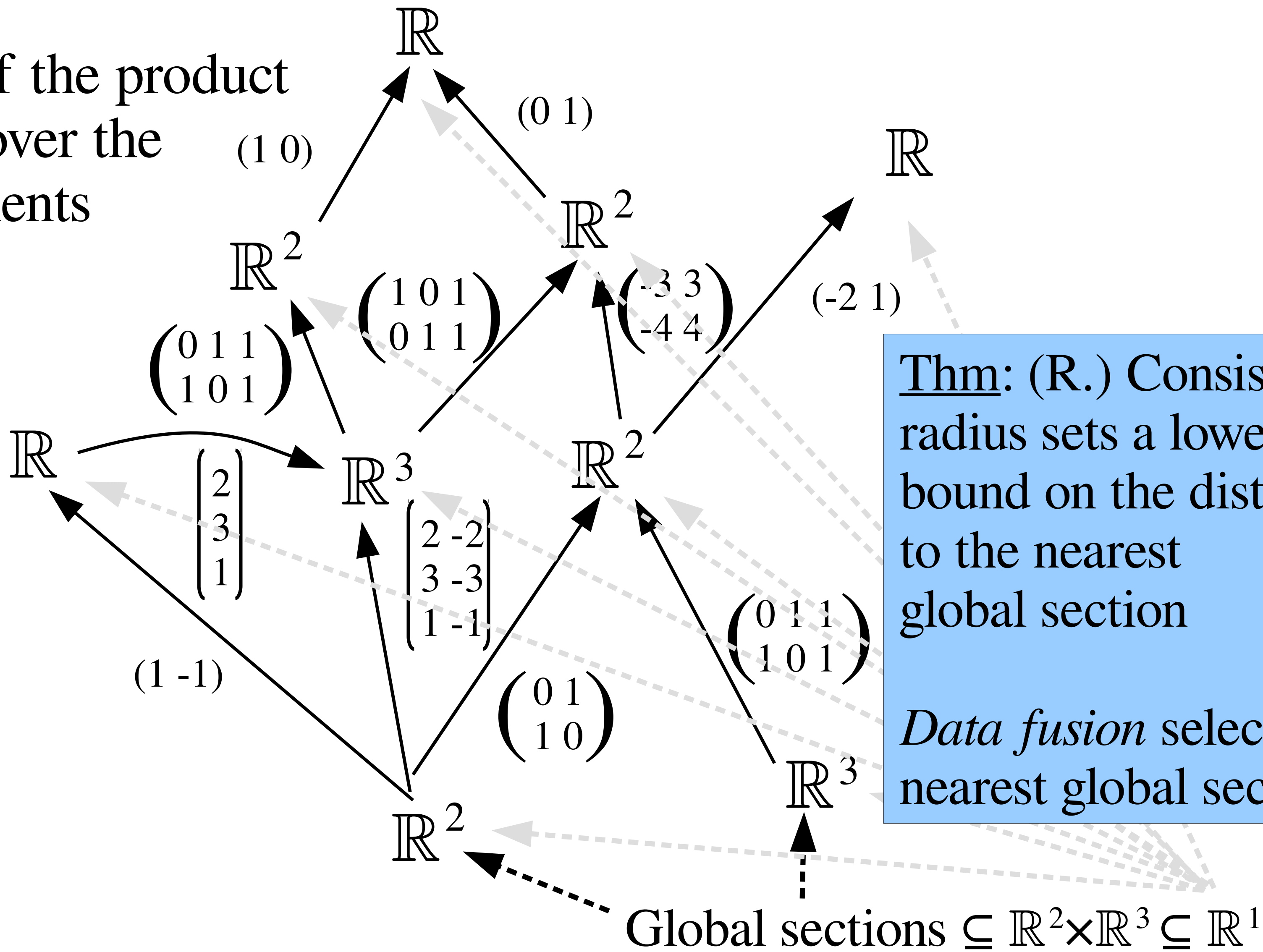


$$\text{MAX} \geq 2\sqrt{14}$$

Note: lots more restrictions to check!

The space of global sections

It's a subset of the product of the stalks over the minimal elements



<u>Thm</u>: (R.) Consistency radius sets a lower bound on the distance to the nearest global section

Data fusion selects the nearest global section

The future

- Computational sheaf theory
  - Small examples can be put together *ad hoc*
  - Larger ones require a software library
- `PySheaf`: a software library for sheaves
  - https://github.com/kb1dds/pysheaf
  - The example is a unit test you can play with!
- Connections to statistical models need to be explored
- Extensive testing on various datasets and scenarios