A categorical approach to high-assurance science

David I. Spivak

dspivak@math.mit.edu
Mathematics Department
Massachusetts Institute of Technology

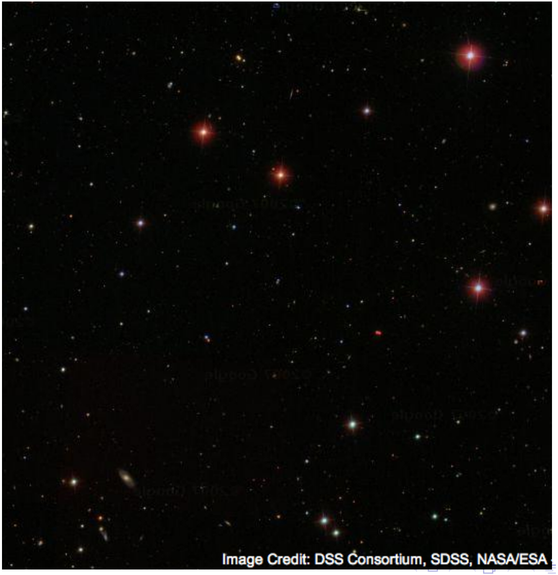
Presented on 2012/06/13
at the Office of Naval Research

Science and map-making

Map-making: The earliest known scientific pursuit.

- Humans have been making maps for at least 8,000 years.
- They made maps of land and maps of the night sky.
- Different people made different observations.
- Comparing these observations and finding commonalities led to
  - increased reliability (peer review),
  - increased scope (division of labor),
  - increased cohesiveness of culture.

First view of the night sky



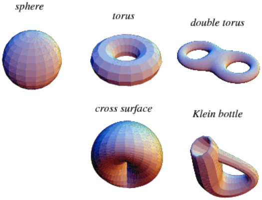
Second view of the night sky: saccade right

Third view of the night sky: saccade down-left

Three views put side by side

Three views put together: one big picture

Manifolds



(Source: Wolfram MathWorld)

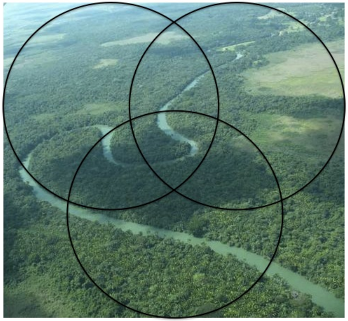
A manifold M is a mathematical shape (a topological space).

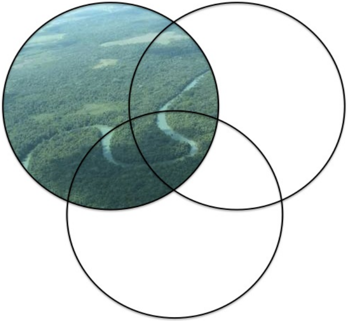
- It can be arbitrarily complex.
- It can have arbitrarily high dimension (e.g. $\dim(M) = 58$).

Two rules set manifolds apart:

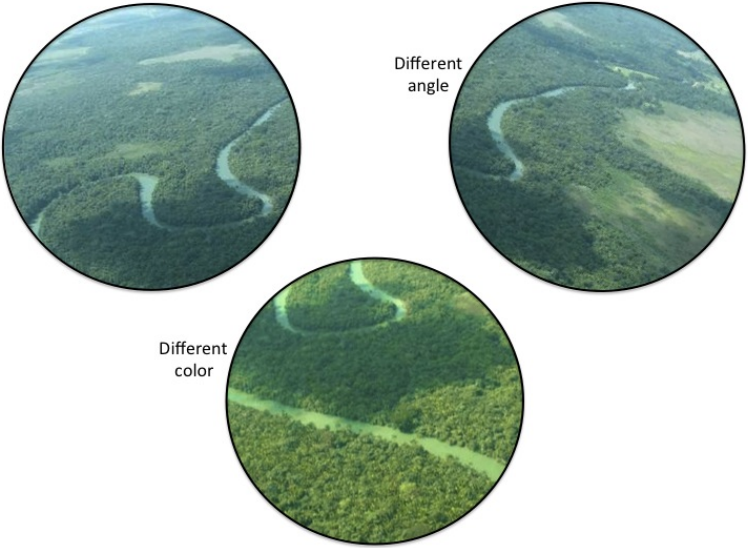
- Locally, a manifold is completely understandable.
  - Each point has a neighborhood that's equivalent to $\mathbb{R}^n$.
- Each set of overlapping local pictures can be meaningfully compared.
  - Overlapping $\mathbb{R}^n$'s are compared by diffeomorphisms.

Overlapping views

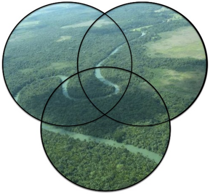




Stitch together isomorphic but unequal overlaps (sheaves)

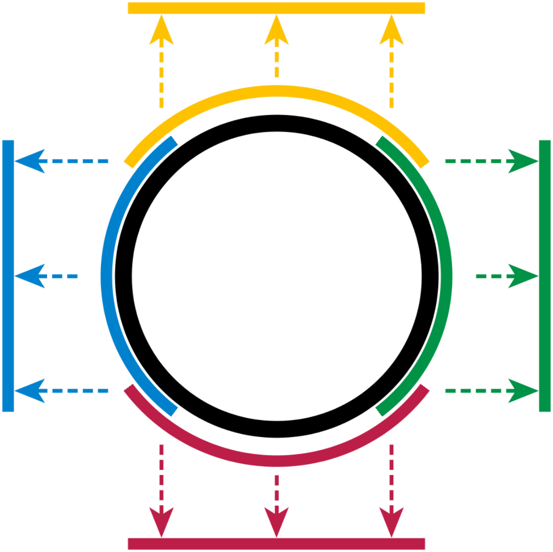


Linked local charts = Atlas



- Each local picture is called a *chart*.
- These charts are linked together by finding equivalent sub-charts.
- Together this system of linked charts forms an *atlas*.
- Crucial feature: controlled linkages
    - We have defined in advance the structure of each local picture.
    - We have defined in advance the structure of comparison on overlaps.

Atlas: Four overlapping views of the black circle



From wikipedia.org.

Expanding on the manifolds idea

- Humans are information processing agents.
- We construct large-scale understanding out of simple local pictures.
  - Worldwide geographical information has been compiled from local data.
  - How are do we put together the mountains of local data into a complete picture?
  - Like a jigsaw puzzle: We look for agreement and fasten along it.
- This basic charts-to-atlas idea underlies the pursuit of knowledge.
- Can we use the charts-to-atlas formalism more broadly?

Individuals as explorers

Imagine that each person is an explorer of his or her world.

- This world is more than spatial, more than visual:
- It includes every kind of information.
- Anything that can be classified, understood, or made sense of.
  - Physics, chemistry, mathematics,
  - fighting, war, logistics,
  - courtesy, relationship, cultures,
  - history, linguistics, psychology.
- Humans have explored a vast informational territory.

Where we are now

Each explorer (person) has made sense of a swath of the world.

- Individual entities (scientists, businesses) often have mature understandings.
- How is this knowledge passed along?
- Information is shared in a weak way:
    - by word of mouth,
    - by imitation,
    - by prose text.
- We need a more robust, rigorous way to share information.

Analogy to software

- To make a working program, one can cobble something together.
- Weakly-typed languages (Perl) are useful for quickly producing individual scripts.
- Strongly-typed languages (Haskell, ML) are much more robust and safe.
  - High-assurance software.
  - Haskell is used in NSA projects (Galois), pharmaceutical companies (Amgen).
  - More scalable, easier to build on, longer life.
- Difference: what is passed
  - strongly-typed languages pass values having ready-made interpretation.
  - weakly-typed languages pass values requiring dynamic interpretation.

Creating high-assurance science

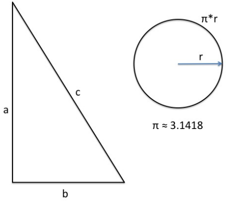
- Individual scientists perform experiments and determine values.
    - For example, the Young's modulus of a material.
    - The throughput of a transportation network.
- How are these results passed from researcher to researcher?
    - In paper publications, as prose text.
    - In talks, as spoken English.
    - By imitation of lab-mates, advisors.
    - These all require human judgment, educated guessing.
- We can do better: strongly typed ideas.

Firm foundations: mathematics and reality

Good science is grounded in two worlds:

- the real world of observation and experiment, and
- the conceptual world of rigorous mathematics.

Example: Geometry and map-making



Pythagorean theorem: $a^2+b^2=c^2$

- Geometry was contemporary with early map-making.
  - Pythagorean theorem known 4,000 years ago.
  - π known to be about 3.1418 by Archimedes in 250 BC.
- Mathematics served as a check for map-making:
  - The data of observation had to conform with geometric principles.
  - Non-conforming maps may have serve temporarily as a heuristic, but they would fall apart under stress or scrutiny.

Good science includes good communication of science

- Today, scientific studies are already firmly grounded in mathematics.
- But science is much more than individual studies.
- It is a network of scientists learning from each other.
- The *communication* of science must be made formal and rigorous.

What is needed from mathematics?

Mathematics can provide:

- A language in which to carefully record all sorts of information.
- A computational toolset with which to manipulate the information.
- A mathematical basis for an atlas of scientific ideas:
  - Each scientist is an explorer, mapping out some territory.
  - We want to rigorously connect these charts into an atlas of human knowledge.

Without mathematics, this would be just a pie-in-the-sky idea.

- The goal of today's talk is to suggest an approach to formalizing it.
- Essential ingredient: connecting databases and category theory.
- This allows us to create an atlas of databases.

Outline:

We want to connect information and mathematics.

1. Discuss what information is, and how we work with it currently.
2. Discuss what category theory is.
3. Show the essential similarity between these subjects.
4. Discuss linkages between disparate viewpoints.
5. Review.

What is information?

- There is plenty of information being produced and used.
- But it is hard to say exactly what information *is*.
- Some sources of information:
  - Dictionaries.
  - Engineer's schematic diagrams.
  - Architect's floor plans.
  - Databases.
- So... what is it?

In formation

What is information?

- Controlled formation!
  - Controlling formation is the same as enforcing order, dispelling chaos.
  - It obviates guessing.
  - It promotes effective reasoning.
  - Information is always in formation.
- Our sources of information:
  - Dictionaries.
  - Engineer's schematic diagrams.
  - Architect's floor plans.
  - Databases.
- Easiest to mathematize: databases.
  - Databases bridge the divide between theory and practice.
  - They include both conceptual layout and on-the-ground facts.

What is a database?

- A database consists of a schema and conforming data.
- Database schema (conceptual layout).
  - A collection of tables.
  - Each table has many columns.
  - The columns refer us from one table to another.
- Database instance (on-the-ground facts).
  - A database instance is a collection of data.
  - Each table is filled with rows of data.
  - All the data is in accordance with the schema.

Example database instance

A family of linked tables:

dog			
ID	**name**	**owner**	**address**
D101	Wally	P34	15 Ash St.
D102	Fido	P46	201 5th Ave.
D104	Buster	P17	27 Spring Ln.

person		
ID	**lastName**	**address**
P17	Jones	27 Spring Ln.
P19	Smith	201 Gladys Ave.
P34	Smith	15 Ash St.
P46	D'Angelo	201 5th Ave.

dogName
ID
Barkie
Buster
Fido
Puppers
Rosie
Samson
Wally

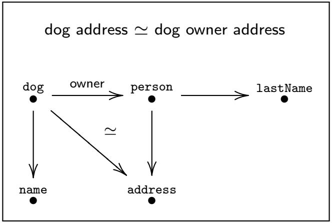
address
ID
15 Ash St.
27 Spring Ln.
201 5th Ave.
201 Gladys Ave.

lastName
ID
Bennet
D'Angelo
Jimenez
Jones
Moran
Smith
Vickers

Database schemas enforce order

- A family of tables is organized by a data architect.
- This organizational structure is called a *schema*.
- A schema specifies precisely:
  - The set of tables and their columns.
  - How the tables interrelate.

A database and its schema



dog			
ID	**name**	**owner**	**address**
D101	Wally	P34	15 Ash St.
D102	Fido	P46	201 5th Ave.
D104	Buster	P17	27 Spring Ln.

person		
ID	**lastName**	**address**
P17	Jones	27 Spring Ln.
P19	Smith	201 Gladys Ave.
P34	Smith	15 Ash St.
P46	D'Angelo	201 5th Ave.

name
ID
Buster
⋮

address
ID
15 Ash St.
⋮

lastName
ID
D'Angelo
⋮

Goal: a mathematical foundation for databases

- The world's information is stored in databases.
- I wanted to find a mathematical basis for databases which:
  - Completely describes schemas, instances, and the relationship between them.
  - Formalizes all typical database operations and querying.
  - Simplifies schema evolution, data migration, and database merging.
  - Links with other information paradigms (RDF and programming languages).
  - Offers new insights and tools.
- The simpler, the better.

What is category theory?

- Since its invention in the early 1940s, category theory has revolutionized math.
- It's like set theory and logic, except less floppy, more principles-based.
- It was invented to build bridges between disparate branches of math by distilling the essence of mathematical structure.
- Original use: connecting topology and algebra.
  - The essence of each was formulated as a category.
  - Rigorous mappings (functors) were established, connected these two categories.
  - These mappings were used to import theorems from algebra as new theorems in topology.

Category theory: branching out

- Category theory naturally fosters connections between disparate fields.
- It has branched out of math and into physics, linguistics, materials science, and biology.
- It has had much success in computer science.
  - Specifically important in the theory of programming languages.
  - The category-theoretic concept of *monads* has vastly extended the reach of functional programming.
- Success comes from simplicity.

Definition of a category I: Constituents

A *category* $\mathcal{C}$ consists of the following constituents:

1. A set $\mathbf{Ob}(\mathcal{C})$, called *the set of objects of* $\mathcal{C}$.
   - I'll denote each object $x \in \mathbf{Ob}(\mathcal{C})$ by $\overset{x}{\bullet}$.
2. A set $\mathbf{Arr}(\mathcal{C})$, called *the set of arrows of* $\mathcal{C}$, and two functions

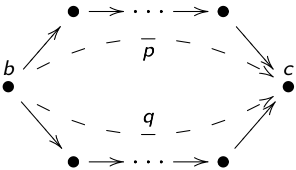
$$src, tgt \colon \mathbf{Arr}(\mathcal{C}) \to \mathbf{Ob}(\mathcal{C}),$$

   assigning to each arrow its *source* and its *target* object, respectively.
   - An arrow $f \in \mathbf{Arr}(\mathcal{C})$ is often written $\overset{x}{\bullet}\xrightarrow{f}\overset{y}{\bullet}$, where $x = src(f), y = tgt(f)$.
   - We define a *path in* $\mathcal{C}$ to be a finite "head-to-tail" sequence of arrows in $\mathcal{C}$, e.g. $\overset{x}{\bullet}\xrightarrow{f}\overset{y}{\bullet}\xrightarrow{g}\overset{z}{\bullet}$.
   - Paths can have length n for any $n \in \mathbb{N}$, including $n = 0$ and $n = 1$.
3. An notion of equivalence for paths, denoted $\simeq$.

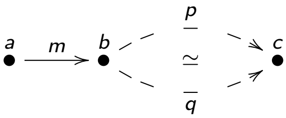
Definition of a category II: Rules

These constituents must satisfy the following requirements:

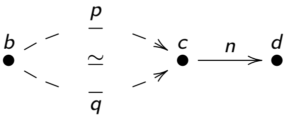
1. If $p \simeq q$ are equivalent paths then the sources agree: $src(p) = src(q)$.
2. If $p \simeq q$ are equivalent paths then the targets agree: $tgt(p) = tgt(q)$.
3. Suppose we have two paths (of any lengths) $b \to c$:



If $p \simeq q$ then for any extensions



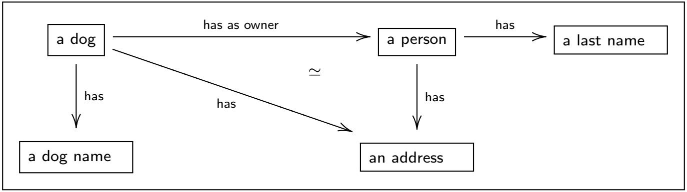
or



$m; p \simeq m; q$ and $p; n \simeq q; n$.

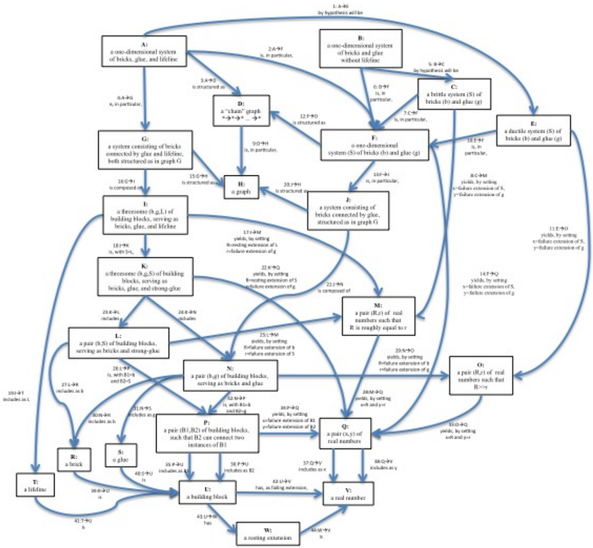
Ologs connect people, databases, and categories

- It turns out that categories and database schemas have the same structure!
- I call the connection between them ologs.



- An olog is both a database schema and a category.

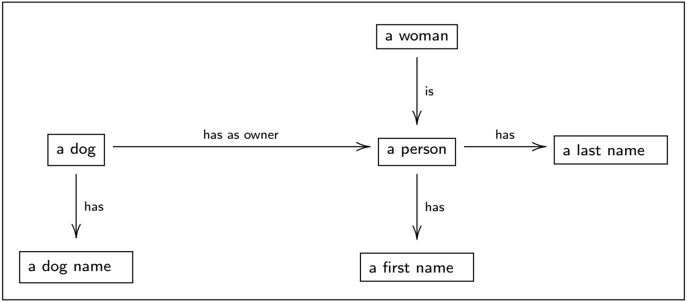
Example: an olog describing hierarchical protein materials



What is an olog?

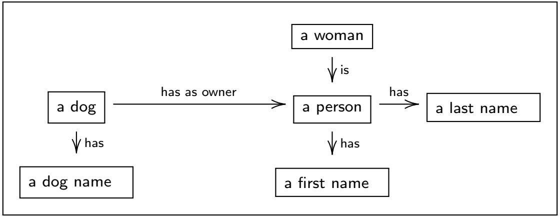
- An olog is a conceptual description of a subject.
- Olog stands for "ontology log"
  - Ontology is the study of what something *is*.
  - "Log" because the study is never complete—always expanding.
- Components of an olog:
  - Labeled boxes,
  - Labeled arrows,
  - Path equivalences.

Ologs are database schemas 1: an example olog



- Boxes are tables
- Arrows are columns.
  - We can predict how many columns the a dog table will have.

Ologs are database schemas 2: database

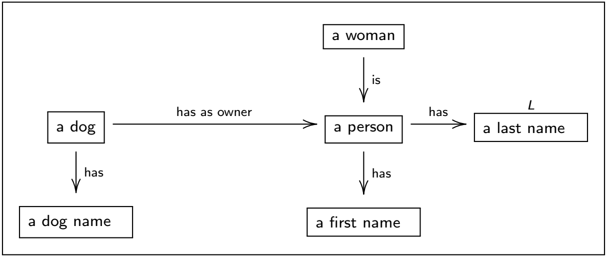


a woman	
ID	**is a person**
W17	P17
W34	P34
W38	P38
W51	P51

a dog		
ID	**has as owner a person**	**has a dog name**
D101	P34	Wally
D102	P46	Fido
D103	P34	Samson
D104	P17	Buster
D106	P19	Rosie

a person		
ID	**has a first name**	**has a last name**
P17	Alice	Jones
P19	Bob	Smith
P34	Barbara	Smith
P38	Sandra	Moran
P46	Jeremy	D'Angelo
P51	Luisa	Jimenez

Leaf tables

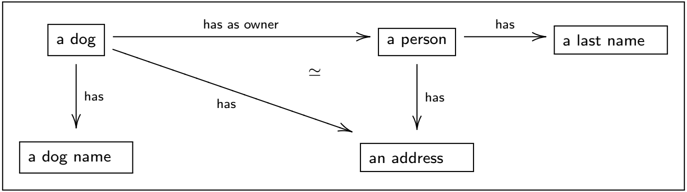


a dog name
ID
Barkie
Buster
Fido
Puppers
Rosie
Samson
Wally

a first name
ID
Alice
Bob
Barbara
Carl
Jeremy
Luisa
Sandra
Thomas

a last name
ID
Bennet
D'Angelo
Jimenez
Jones
Moran
Smith
Vickers

Equivalent paths require equivalent data

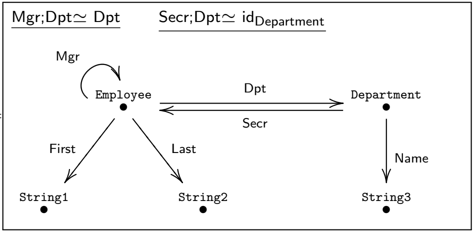


dog			
ID	**name**	**owner**	**address**
D101	Wally	P34	15 Ash St.
D102	Fido	P46	201 5th Ave.
D104	Buster	P17	27 Spring Ln.

person		
ID	**lastName**	**address**
P17	Jones	27 Spring Ln.
P19	Smith	201 Gladys Ave.
P34	Smith	15 Ash St.
P46	D'Angelo	201 5th Ave.

Another example of path equivalences

$\mathcal{C} :=$



Employee				
Id	**First**	**Last**	**Mgr**	**Dpt**
101	David	Hilbert	103	q10
102	Bertrand	Russell	102	x02
103	Alan	Turing	103	q10

Department		
Id	**Name**	**Secr**
q10	Sales	101
x02	Production	102

String
Id
a
b
⋮
z
aa
ab
⋮

Ologs bridge the divide

- Each olog is authored by an individual or group, about a subject.
  - The olog idea can be understood by ordinary people.
  - No database theory or category theory necessary.
- Ologs are both databases and categories, in disguise.
  - Ologs are database schemas; we can fill them with relevant data.
  - Ologs are categories; mathematics can be brought to bear.
- I will use the following words interchangeably:
  - Olog,
  - Database schema,
  - Category.

Relating different families of tables

- An olog is the mathematical structure underlying a database.
  - Each database is a specific layout for a whole family of tables.
- We want to link different ologs together.
- Example 1: Banks
  - Each bank has its own database schema.
  - No two banks structure their tables in exactly the same way.
  - The Federal Reserve wants to understand the whole picture.
- Example 2: Formal network of science.
- How can mathematics help?

Linking databases together

- Forming a coherent whole.
  - Different scientists or different banks may structure their data differently.
  - If they are studying the same subject, links should exist.
  - We want to stitch differently-structured schemas together.
  - Connecting different schemas is the same as connecting different categories.
- Category theory was designed specifically for this.
- Next we will discuss the links between categories, called *functors*.

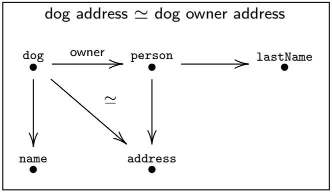
Functors: mappings between categories

- One way to think of a category is as a directed graph, where certain paths have been declared equivalent.
- A functor is a graph-mapping that is required to respect equivalence of paths.
- **Definition**: A functor $F\colon \mathcal{C} \to \mathcal{D}$ consists of
  - a function $\mathbf{Ob}(\mathcal{C}) \to \mathbf{Ob}(\mathcal{D})$ and
  - a function $\mathbf{Arr}(\mathcal{C}) \to \mathbf{Path}(\mathcal{D})$,

  such that F
  - respects sources and targets,
  - respects equivalences of paths.

Backing up: a database instance is a functor!

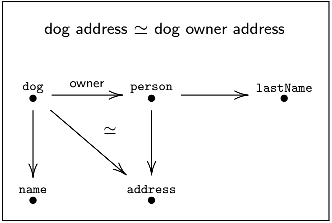
- A database schema (layout of tables) is simply a category $\mathcal{C}$.

$\mathcal{C} :=$ 

- There is a category **Set** of sets and functions.
- A functor $I\colon \mathcal{C} \to \mathbf{Set}$ assigns:
  - to each object $c \in \mathbf{Ob}(\mathcal{C})$ a set $I(c)$,
  - to each arrow $h\colon c \to d$ in $\mathcal{C}$ a function $I(h)\colon I(c) \to I(d)$,
  - such that all path equivalences are respected.
- In other words, a functor $I\colon \mathcal{C} \to \mathbf{Set}$ is a database instance on $\mathcal{C}$; i.e. it is a way to fill $\mathcal{C}$ with compatible data.

Example

$\mathcal{C} :=$



We can represent a functor

$$I : \mathcal{C} \to \textbf{Set}$$

as follows:

dog			
ID	**name**	**owner**	**address**
D101	Wally	P34	15 Ash St.
D102	Fido	P46	201 5th Ave.
D104	Buster	P17	27 Spring Ln.

person		
ID	**lastName**	**address**
P17	Jones	27 Spring Ln.
P19	Smith	201 Gladys Ave.
P34	Smith	15 Ash St.
P46	D'Angelo	201 5th Ave.

name
ID
Buster
⋮

address
ID
15 Ash St.
⋮

lastName
ID
D'Angelo
⋮

Changes in schema

- Suppose in our modeling of a given subject, we evolve from schema $\mathcal{C}$ to schema $\mathcal{D}$.
- We should find a functorial connection between them.
- Over time we may have something like

$$\mathcal{C} = \mathcal{C}_0 \xrightarrow{F_0} \mathcal{C}_1 \xrightarrow{F_1} \cdots \xrightarrow{F_n} \mathcal{C}_n = \mathcal{D}$$

- We want to be able to migrate data from $\mathcal{C}$ to $\mathcal{D}$ and vice versa.
- We want to be able to migrate queries against $\mathcal{C}$ to queries against $\mathcal{D}$ and vice versa.
- And we want this all to work as expected.

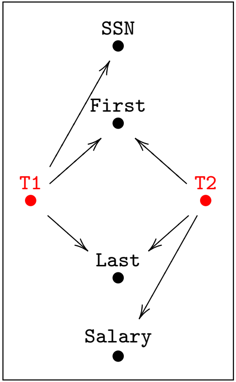
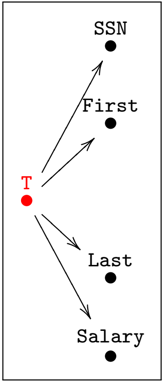
Functorial data migration for CT experts

- For any schema (category) $\mathcal{C}$, we have the category $\mathcal{C}$–**Set** of set-valued functors $I\colon \mathcal{C} \to \mathbf{Set}$ and natural transformations. These are the instances of the database.
- A functor $F\colon \mathcal{C} \to \mathcal{D}$ serves as a translation between schemas.
- Composition with F induces a functor $\Delta_F\colon \mathcal{D}\text{–}\mathbf{Set} \to \mathcal{C}\text{–}\mathbf{Set}$,

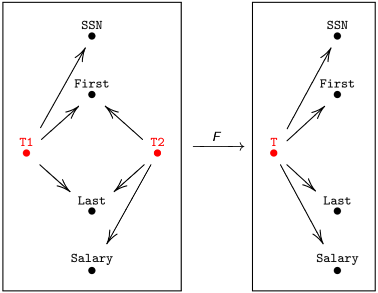
$$\mathcal{C} \xrightarrow{F} \mathcal{D} \xrightarrow{I} \mathbf{Set}.$$

- The functor Δ_F migrates data from $\mathcal{D}$ back to $\mathcal{C}$.
- It has two adjoints $\Sigma_F\colon \mathcal{C}\text{–}\mathbf{Set} \to \mathcal{D}\text{–}\mathbf{Set}$ and $\Pi_F\colon \mathcal{C}\text{–}\mathbf{Set} \to \mathcal{D}\text{–}\mathbf{Set}$.

Uses of functorial data migration 1: Translation F

$\mathcal{C} :=$  $\xrightarrow{F}$  $=: \mathcal{D}$

Uses of functorial data migration 2: Projection via Δ_F

$\mathcal{C} :=$  $=: \mathcal{D}$

$J\colon \mathcal{D} \to$ **Set**:

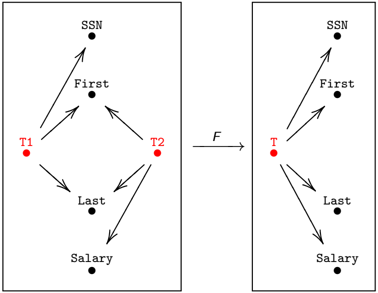
T				
ID	**SSN**	**First**	**Last**	**Salary**
XF667	115-234	Bob	Smith	$250
XF891	122-988	Sue	Smith	$300
XF221	198-877	Alice	Jones	$100

$\Delta_F(J)\colon \mathcal{C} \to$ **Set**:

T1			
ID	**SSN**	**First**	**Last**
XF667T1	115-234	Bob	Smith
XF891T1	122-988	Sue	Smith
XF221T1	198-877	Alice	Jones

T2			
ID	**First**	**Last**	**Salary**
XF667T2	Bob	Smith	$250
XF891T2	Sue	Smith	$300
XF221T2	Alice	Jones	$100

Uses of functorial data migration 3: Joins via Π_F

$\mathcal{C} :=$  $=: \mathcal{D}$

$I\colon \mathcal{C} \to$ **Set**:

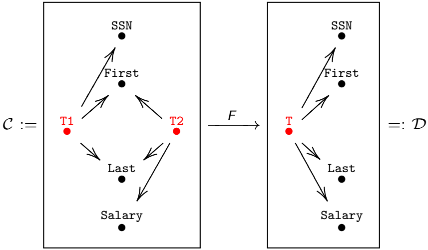
T1			
ID	**SSN**	**First**	**Last**
T1-001	115-234	Bob	Smith
T1-002	122-988	Sue	Smith
T1-003	198-877	Alice	Jones

T2			
ID	**First**	**Last**	**Salary**
T2-A101	Alice	Jones	\$100
T2-A102	Sam	Miller	\$150
T2-A104	Sue	Smith	\$300
T2-A110	Carl	Pratt	\$200

$\Pi_F(I)\colon \mathcal{D} \to$ **Set**:

T				
ID	**SSN**	**First**	**Last**	**Salary**
T1-002T2-A104	122-988	Sue	Smith	\$300
T1-003T2-A101	198-877	Alice	Jones	\$100

Uses of functorial data migration 4: Unions via Σ_F



$I\colon \mathcal{C} \to$ **Set**:

T1			
ID	**SSN**	**First**	**Last**
T1-001	115-234	Bob	Smith
T1-002	122-988	Sue	Smith
T1-003	198-877	Alice	Jones

T2			
ID	**First**	**Last**	**Salary**
T2-A101	Alice	Jones	\$100
T2-A102	Sam	Miller	\$150
T2-A104	Sue	Smith	\$300
T2-A110	Carl	Pratt	\$200

$\Sigma_F(I)\colon \mathcal{D} \to$ **Set**:

T				
ID	**SSN**	**First**	**Last**	**Salary**
T1-001	115-234	Bob	Smith	T1-001.Salary
T1-002	122-988	Sue	Smith	T1-002.Salary
T1-003	198-877	Alice	Jones	T1-003.Salary
T2-A101	T2-A101.SSN	Alice	Jones	\$100
T2-A102	T2-A102.SSN	Sam	Miller	\$150
T2-A104	T2-A104.SSN	Sue	Smith	\$300
T2-A110	T2-A110.SSN	Carl	Pratt	\$200

Category theory provides a foundation for information

- We can formulate an understanding of any topic using an olog.
- The olog can then be filled with conforming data.
  - Categories capture the structure of databases completely.
  - The olog structure points out variables that are critical to our understanding.
  - All in an intuitive yet rigorous way.
- All typical manipulations of the data are grounded in pure math.
  - Simple data shuffling (projections, unions, joins, warehousing, etc.)
  - Aggregations (sums, counts, averages).
  - Curve fitting, parameter estimation, classifying results.
  - Schema evolution, data migration, merging.

Category theory has 70 years worth of useful theorems

- There is perfect correspondence between database and categories.
- Theorems about categories are theorems about databases.
- We can thus *prove* things about how information works.

Ask me later about:

There wasn't space to fit in the following neat examples:

- Connection to RDF and semi-structured data via the Grothendieck construction.
  - SPARQL graph-pattern queries become topological lifting problems.
- A normal form for queries using a purely categorical theorem.
- Aggregation functions and hierarchical categories.
- Probabilistic ologs with Bayesian updating.
- Moggi-style use of monads in categorical databases.

Advantages of a mathematical foundation

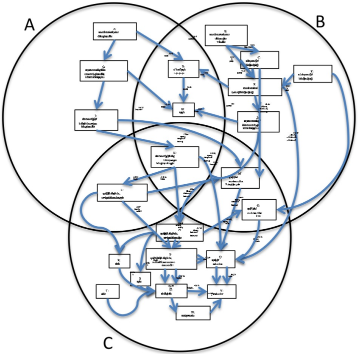
- A unified language for information science. (Interoperable)
- Easily adaptable to different platforms. (Portable)
- Database operations are founded and certifiable. (Rigorous)
  - We can prove things about the results of a query or data transformation before performing it.
  - Mathematics does not fail under pressure.
- The underlying mathematics is the same at all size scales. (Scalable)

Certified science

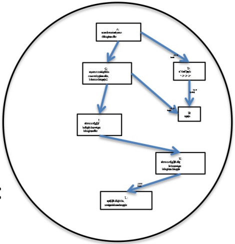
Envision this process:

- A scientific topic is formulated as an olog.
- A proof assistant (Coq) is used to transform the mathematical toolset described above into a certified program.
    - This program collects the data and performs the manipulations.
    - Parameter estimation, curve fitting, statistical diagnoses are all provable.
    - All scientific claims are proven, and others can investigate the evidence at any scale.
    - Information is more readily shared and processed.
    - At each step, a mathematical proof of correctness (certificate) is produced by Coq.
- Different topics can be fused together to create a robust network of human understanding.

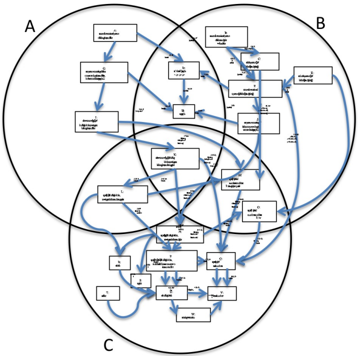
Network of scientists 1: overlapping understanding

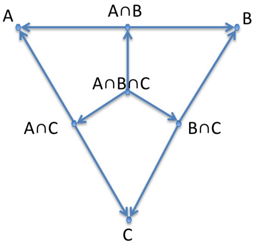


Scientist A's research topic:

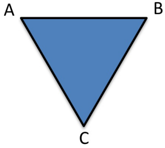


Network of scientists 2: encoding interaction groups



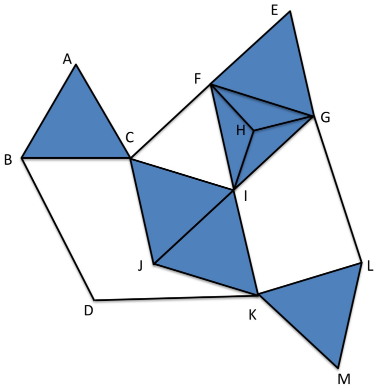


Abstraction:



Network of scientists 3: simplicial complex

A network of ologs, a network of scientific understanding.



This whole network can be queried, with provenance plainly evident.

A firm foundation for communication

We should study the communication process.

- Ground in the real world of observation and experiment.
  - How is information stored, processed, transferred currently?
  - This has not caught on as a mathematical pursuit.
- Check using the rigor of mathematics.
  - Find tried-and-true mathematical structures on which to base the study.
  - Practitioners are generally uninterested in this.
- A healthy combination will have profound affect.
  - Symbiosis between math and physics benefited both fields.
  - Could such a relationship exist between mathematics and information?

My growing network

- Professors in other fields at MIT
  - Markus Buehler, CEE
  - Adam Chlipala, CSAIL
- Mathematics postdocs and professors
  - Scott Morrison (UC Berkeley)
  - Nat Stapleton (MIT)
  - Mathieu Anel (UQAM)
  - David Gepner (Universität Regensburg)
  - Steve Awodey (CMU)
  - Jack Morava (JHU)
- Industry
  - Dave Balaban (VP at Amgen)
  - Peter Gates (Johnson and Johnson)
  - Rich Haney (GlaxoSmithKline)
  - Carlo Curino (Yahoo! Research)
  - Allen Brown (Microsoft Research)

Summary of the talk

- We need to improve our ability to communicate rigorously about complex subjects.
  - Transferring knowledge from one group to another is difficult.
  - It cannot be left to human guessing and ad-hoc interpretation.
  - We need to have available a high-assurance framework for communication.
- Ologs and category theory provide such a framework.
- If broadly adopted, this could have profound impacts on science.
  - A formal strategy for stitching local databases into an atlas of science.
  - Unify each field in vocabulary and agenda.
  - Improve data acquisition strategies.
  - Allow multi-lab data analysis.
- Scientific communication will surely benefit from an infusion of mathematics.

Thank you

Thanks for inviting me to speak!