3. Conceptual Frameworks

Conceptual Frameworks

We use data to represent the world. But we also:

- describe the world using **language**
- represent it by building **physical models**

Common thread: **representation** (an object standing for another, being used in its stead in order to indirectly engage with the object being represented).

On one hand: “the map is not the territory”, but we do not need much effort to use the map to navigate the territory.

The transition from **representation** to **represented** can be seamless, which is risky: **it is easy to mistake the data/analytical results for the real world**.

Conceptual Frameworks

Best protection: thought out and explicitly described **conceptual framework**

- a **specification** of which parts of the world are being represented
- **how** they are represented
- the **nature of the relationship** between the represented and the representing
- **appropriate** and **rigorous strategies** for applying the results of the analysis that is carried out in this representational framework

It could be built from scratch for each new project, but there are **modeling frameworks** that are broadly applicable to many different phenomena, which can be moulded to fit specific instances.

Three Modeling Strategies

There are 3 main (not mutually exclusive) **modeling strategies** that can be used to guide the specification of a phenomenon or domain:

- **mathematical** modeling
- **computer** modeling
- **systems** modeling

The first two have their own mathematical/digital (logical) worlds, distinct from the tangible, physical world studied by chemists, biologists, and so on:

- used to describe real-world phenomena by **drawing parallels** between the properties of objects in different worlds and reasoning via these parallels.

Three Modeling Strategies

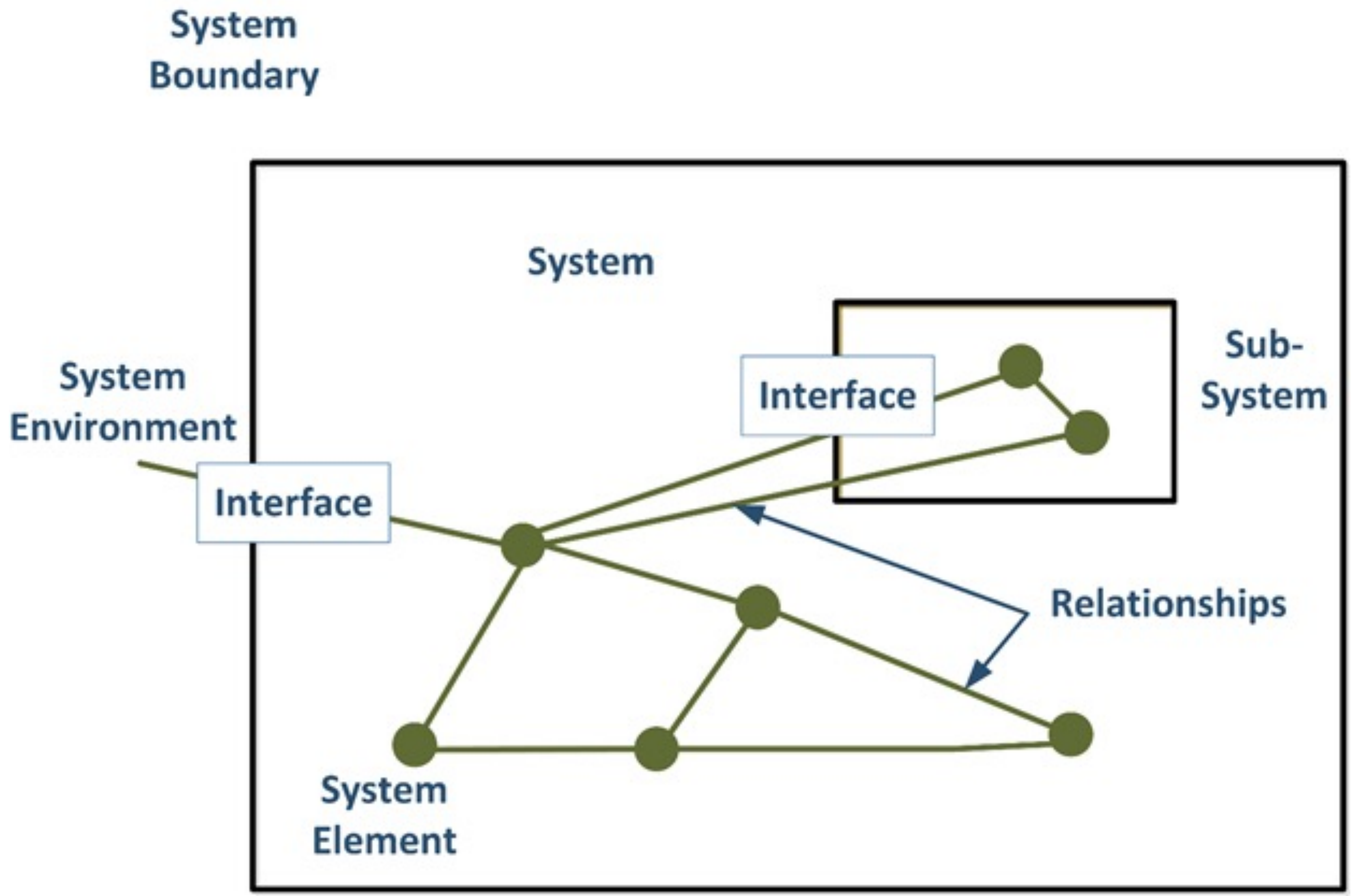
General Systems Theory describes **disparate** natural phenomena using a **common conceptual framework**, all as systems of interacting objects.

When presented with a new situation, we ask ourselves:

- which objects seem most relevant in the system behaviours of interest?
- what are the properties of these objects?
- what are the behaviours (or actions) of these objects?
- what are the relationships between these objects?
- how do the relationships between objects influence their properties and behaviours?

Goals: **understand the system** and **relevant behaviours,** develop consistent **shared understanding**, inform **data collection, guide data interpretation**.

[Deepak, General Systems Theory, Wikimedia]



Information Gathering

Achieving **contextual understanding** of a dataset is crucial.

Concretely, how does this understanding come about?

It can be reached through:

- **field trips**
- interviews with **subject matter experts** (SMEs)
- **readings/viewings**
- **data exploration** (even just **trying to obtain** or gain access to the data can prove a major pain), etc.

Information Gathering

Clients or stakeholders are **not uniform** entities – client data specialists and SMEs may **resent the involvement** of analysts (external and/or internal).

Information gathering provides analysts the opportunity to show that everyone is pulling in the same direction, by:

- asking **meaningful** questions
- taking a **genuine interest** in the SMEs'/clients' experiences
- acknowledging everyone's ability to contribute

A little tact goes a long way when it comes to information gathering.

Thinking in Systems Terms

A **system** is made up of **objects** with **properties** that can change over time.

Within the system, there are **actions** and **evolving properties**, i.e., **processes**.

We understand how various aspects of the world interact with one another by **carving chunks** corresponding to the aspects and define their boundaries.

Working with other intelligences requires a **shared understanding** of what is being studied.

Objects themselves have various properties.

Thinking in Systems Terms

Natural processes generate/destroy objects, and change the properties of these objects over time.

We **observe**, **quantify**, and **record** values of these properties at particular points in time.

Observations are used to **capture the underlying reality** to an acceptable degree of **accuracy** and **error**, but … **even the best system model only ever provides an approximation of the situation under analysis.**

With luck, experience, foresight, these approximations might be **valid**.

Identifying Gaps in Knowledge

A **gap in knowledge** is identified when we realize that what we thought we knew about a system proves **incomplete** (or blatantly false).

Causes:

- naïveté *vis-à-vis* the situation being modeled
- nature of the project under consideration

With **too many moving parts**, **unrealistic objectives**, **distance from pipeline**, knowledge gaps cannot be avoided (although they also occur with small, well-organized, easily contained projects).

Identifying Gaps in Knowledge

Knowledge gaps might occur **repeatedly**, at any moment in the process:

- data **cleaning**
- data **consolidation**
- data **analysis**
- even during **communication of the results** (!)

When faced with a knowledge gap, **be flexible**:

- **go back**
- **ask questions**
- **modify the system representation** as often as is necessary

It is preferable to catch these gaps early on in the process (obviously).

Conceptual Models

Conceptual models are built using methodical investigation tools:

- **diagrams**
- structured **interviews**
- structured **descriptions,** etc.

Data scientists should beware **implicit conceptual models** (knowledge gaps).

It is preferable to err on the side of “too much conceptual modeling”, but remember that “every model is wrong; some models are useful” [G.E. Box].

It is OK to build better models in an iterative manner.

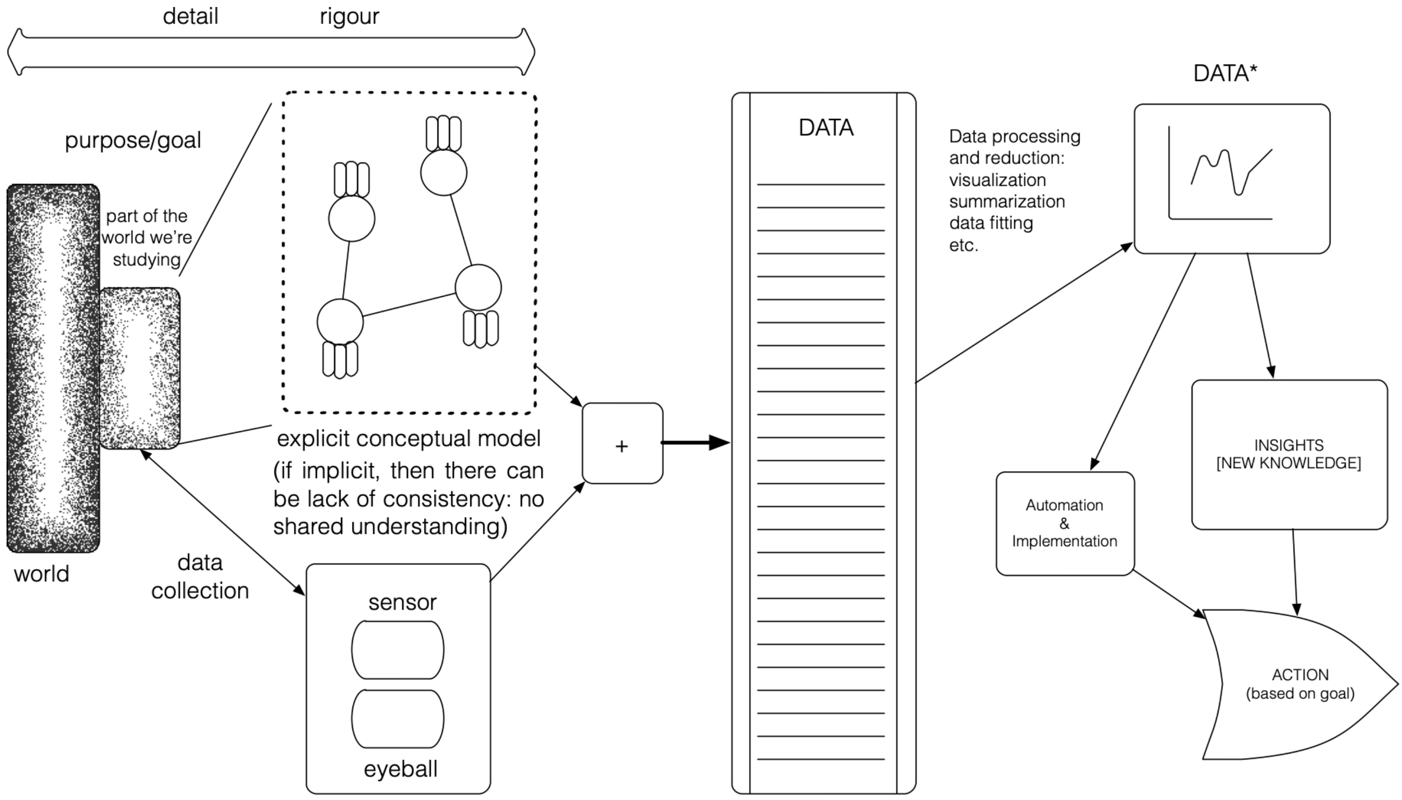
Conceptual Models

Conceptual model

- are not implemented as a scale-model or computer code
- exist only conceptually, often in the form of a diagram/verbal description of a system – boxes and arrows, mind maps, lists, definitions

Focus is on:

- **possible states** (not specific behaviour)
- object types, not specific instances; the goal is **abstraction**

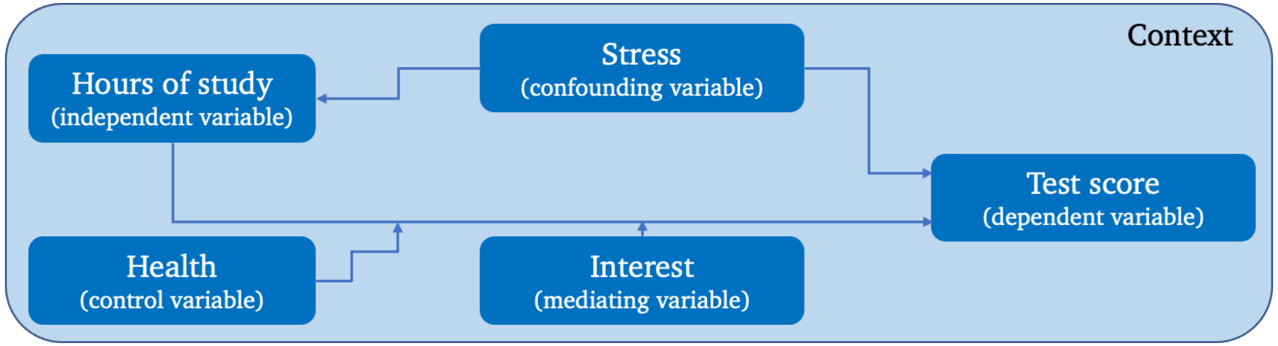


Conceptual Models

In practice, we must first select a system for the task at hand, then generate a conceptual model that encompasses:

- **relevant** and **key objects** (abstract or concrete);
- **properties** of these objects, and their values;
- **relationships between objects** (part-whole, is-a, object-specific, one-to-many), and
- **relationships between properties** across instances of an object type.

A simplistic example describing a supposed relationship between a **presumed cause** (hours of study) and a **presumed effect** (test score).



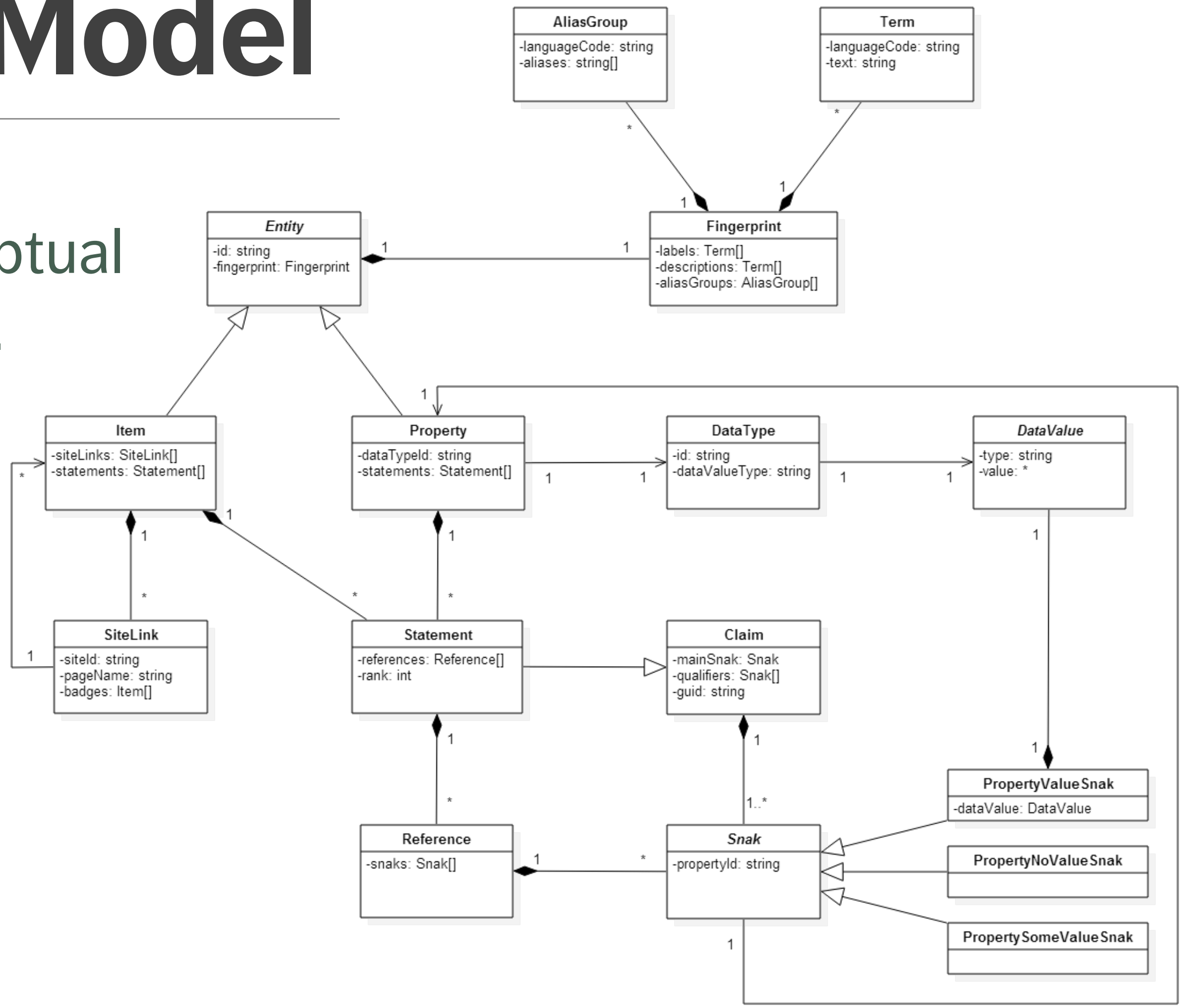
Formal Conceptual Model

Conceptual modeling turn implicit conceptual models into **explicit** and tangible models.

It provides opportunities to examine and explore ideas and assumptions.

Various efforts have been made to **formalize** conceptual modelling:

- UML (Universal Modelling Language)
- Entity Relationship (ER) Models



Relating Data to the System

Is the collected and analyzed data **useful to understand the system**? This question can best be answered if we understand:

- **how** the data is collected
- the **approximate nature** of both data and system
- what the data represents (observations and features)

Is the **combination of system and data sufficient** to understand the situation under consideration? Difficult to answer in practice.

If the data, the system, and the real world are **out of alignment**, any data insight drawn from modeling and analysis might ultimately prove useless.

Cognitive Biases

Cognitive biases have an impact on how we construct models and look for patterns in the data:

- **Anchoring Bias** causes us to rely too heavily on the first piece of information we are given about a topic
- **Availability Heuristic** describes our tendency to use information that comes to mind quickly and easily when making decisions about the future
- **Bandwagon Effect** refers to our habit of adopting certain behaviours or beliefs because many others do the same
- **Choice-Supporting Bias** causes us to view our actions in a positive light, even if they are flawed
- **Clustering Illusion** refers to our tendency to see patterns in random events
- **Confirmation Bias** describes our tendency to notice, focus on, and give greater credence to evidence that fits with our existing beliefs
- **Conservation Bias** occurs when we favour prior evidence over new information
- **Ostrich Effect** describes how people often avoid negative information, including feedback that helps them monitor their goal progress

Cognitive Biases

- **Outcome Bias** refers to judging a decision on the outcome, rather than on why it was made
- **Overconfidence** causes us to take greater risks in our daily lives
- **Pro-innovation Bias** occurs when proponents of a technology overvalue its usefulness and undervalue its limitations
- **Recency Bias** occurs when we favour new information over prior evidence
- **Salience Bias** describes our tendency to focus on items or information that are more noteworthy while ignoring those that do not grab our attention
- **Survivorship Bias** is a cognitive shortcut that occurs when a visible successful subgroup is mistaken as an entire group
- **Zero-Risk Bias** relates to our preference for absolute certainty

Other biases:

- base rate fallacy, bounded rationality, category size bias, commitment bias, Dunning-Kruger effect, framing effect, hot-hand fallacy, IKEA effect, illusion of explanatory depth, illusion of validity, illusory correlations, look-elsewhere effect, optimism effect, planning fallacy, response bias, selective perception, etc.

Suggested Reading

Conceptual Frameworks

Data Understanding, Data Analysis, Data Science
Data Science Basics

Conceptual Frameworks for Data Work

- Three Modeling Strategies
- Information Gathering
- Cognitive Biases

Exercises

Conceptual Frameworks

1. Consider the following situation: you are away on business and you forgot to hand in a very important (and urgently required) architectural drawing to your supervisor before leaving. Your office will send an intern to pick it up in your living space. How would you explain to them, by phone, how to find the document? If the intern has previously been in your living space, if their living space is comparable to yours, or if your spouse is at home, the process may be sped up considerably, but with somebody for whom the space is new (or someone with a visual impairment, say), it is easy to see how things could get complicated. Time is of the essence – you and the intern need to get the job done **correctly** as **quickly as possible**. What is your strategy?

2. Translate the cognitive biases to analytical contexts. What cognitive biases are you, your team, and your organization most susceptible to? Least?