# Data Analysis & ERD

Markku Reunanen / Aalto

Data Analysis

- Often done already earlier than this
- Figuring out
  - Variables
  - Their types, value ranges and accuracy
  - Structure of the data
- Facilitates better understanding of the subject
- Works toward the implementation
- Closely related to object-oriented programming

Entity-Relationship Diagram

- Or just ERD

- A boundary object of a kind
  - Programmers and database designers use them for the implementation
  - Also designers and data analysts can read and make them
  - Facilitates discussion and shared understanding

- Data broken into Entitites, Attributes and their Relationships
  - *Entities* are higher-order data objects
  - *Attributes* are their atomic lower-order values
  - *Relationships* connect entities and can have names too

ERD 1
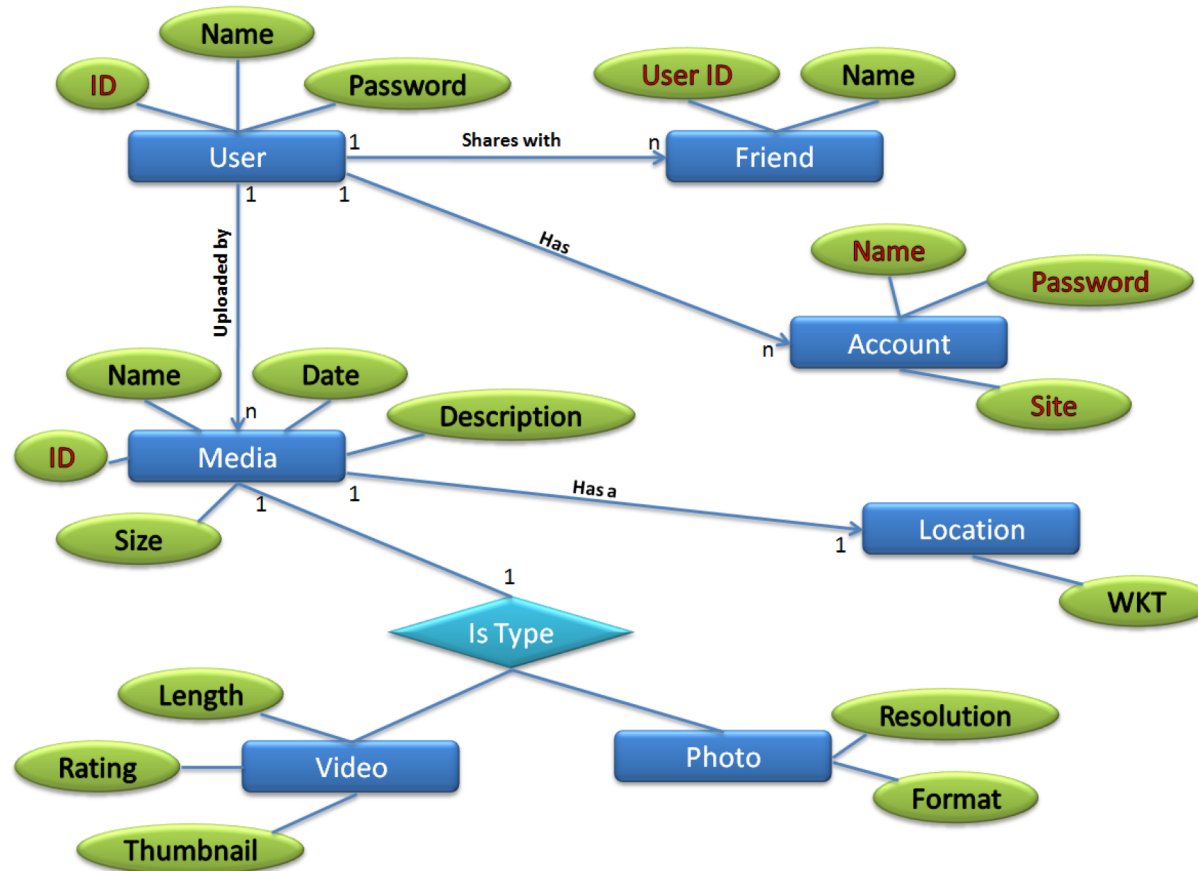
ERD 2
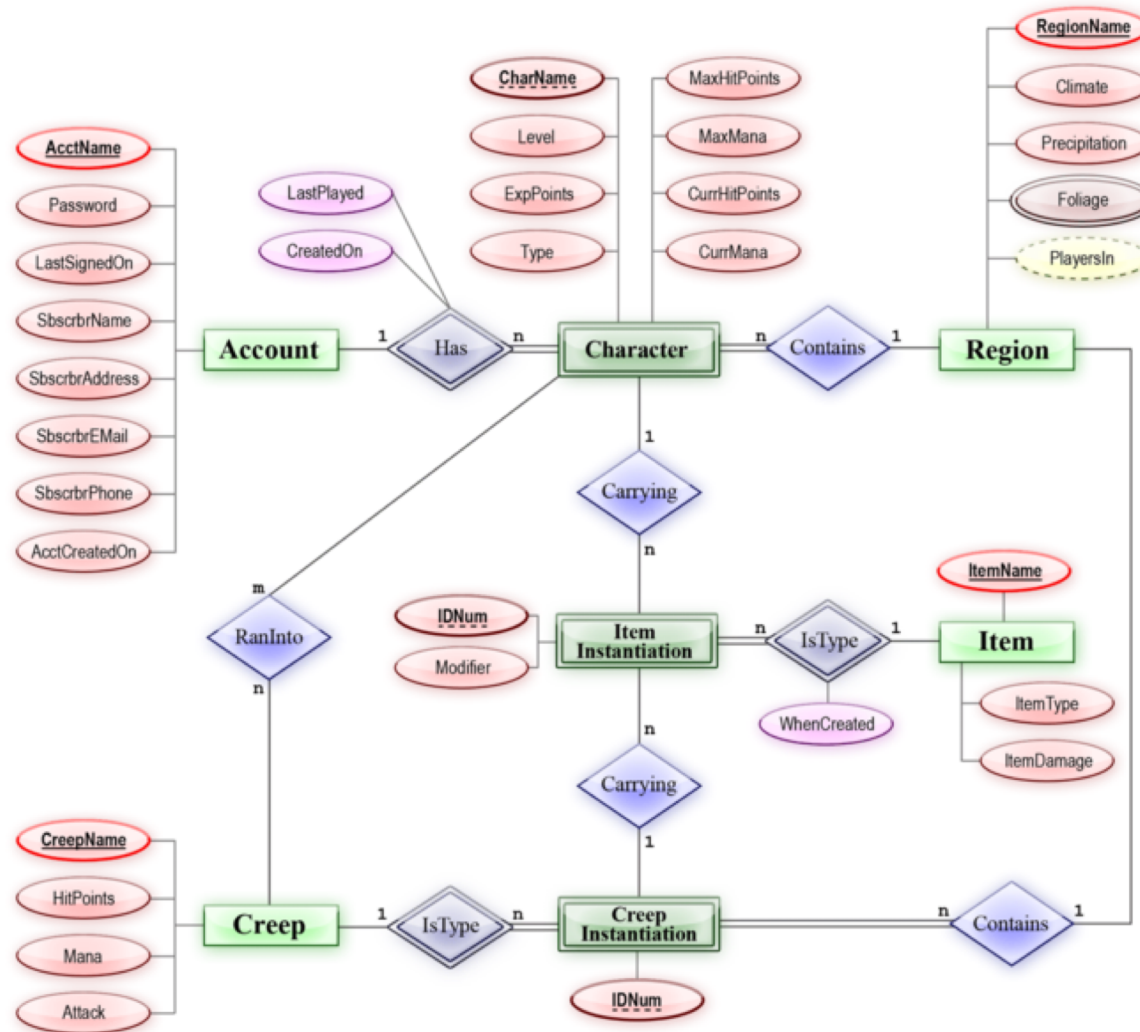
More on ERDs

- Relationships also include *cardinalities*
  - 1..7
  - 2
  - * (any number, at times also n)
  - 2..*
  - They go both ways
  - Cardinality makes plurals unnecessary in entity names (Student, not StudentS)

Some tips

- Start with a flat list and discover its internal structure
- Too complex attribute => make it an entity with multiple attributes
- Too simple entity => make it an attribute
- Many things need an *id*: important for the implementation but not for conceptual thinking
- There is usually more than one solution
- What purpose is this for?

Let's try together first

- An audio CD
  - What's the data? – need to draw the line somewhere
  - What's the structure?
  - Let's make an ERD
- Next: your own data!