



Informatics and Informaticists

Jay Greenfield, PhD

Member of the Data Documentation Initiative (DDI)
Data Description Moving Forward Project



Acknowledgements

- The DDI Moving Forward Data Description Team began working on the structure of a simple data description to cover archival data storage formats and to provide an extension base for describing more complex data description (i.e. databases)
- A shift occurred during the Minneapolis Sprint (2015-05-25)
- There it seemed within reach that at long last we might be able to account for and relate many different types of databases in a single model
- Currently the team is working to prove the Minneapolis vision by using the model presented here to describe several data management solutions

Team Members

Name	Contact	Role
Steve McEachern	steven.mceachern@anu.edu.au	Team Leader
Achim Wackerow	joachim.wackerow@gesis.org	Modeller
Dan Gillman	gillman.daniel@bls.gov	Team member
Larry Hoyle	larryhoyle@ku.edu	Team member
Ørnulf Risnes	ornulf.risnes@nsd.uib.no	Team member
Barry Radler	bradler@wisc.edu	Team member
Justin Lynch	justin.lynch@abs.gov.au	Team member
Jay Greenfield	greenfield_jay@bah.com	Team member
Chris Seymour	chris.seymour@stats.govt.nz	Team member
Wendy Thomas	wlt@umn.edu	Team member

A data model is an information object in a chain of information objects that begins with real world events and things

In this chain events and things are observed and observation yields information

Data models are used to understand this information

Information and information objects like data models and data stores are not new

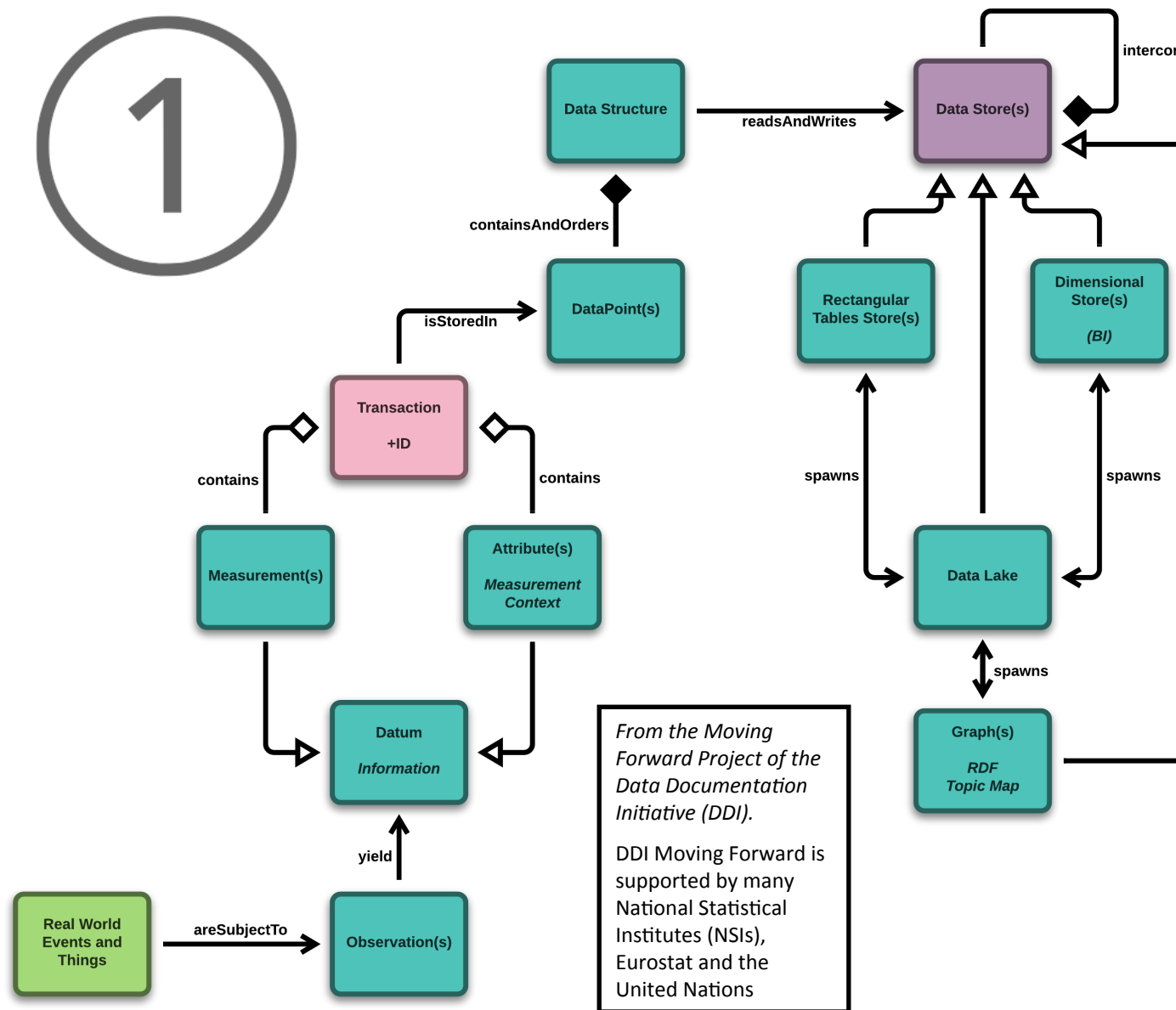
- Books have been around for a while
- So have documents and tables
- However, as a consequence of computers, new types of information objects like dimensional stores in the late twentieth century and now data lakes in the twenty-first century have sprung into being

As a consequence, we now find ourselves in the position of having a new story to tell about information

This story is called *informatics* and the storytellers are called *informaticists*

One surprise in this story are transactions. Transactions preserve context and enable us to think about and understand events and things as a whole

- Context might include *metadata* like a question asked or the kind of cuff used to take a blood measure
- Context might include *paradata* like how long it took to answer a question or the condition of a person during a blood pressure reading



Another surprise is that with the emergence of new types of data models, informaticists can now look at events and things from many *perspectives*

Graphs (RDF) and topic maps (Google) are also data stores

Today, to gain perspective, we don't limit ourselves to a single model

Instead we build models and stores as needed

- Today, increasingly, we build models *downstream*
- Also, one model like a data lake can lead to others like tables and graphs and vice versa
- *Data exploration* – again facilitated by computers – leads us to tweak data structures and build new data stores
- Finally, data stores themselves become nodes in data graphs that we traverse and, in the process, see events and things from multiple *connected* perspectives

Through the myriad of data models, informaticists turn information into knowledge

- We are approaching a time when anyone who is curious can be an informaticist because of the emergence of computers and tools
- The informaticist need not be technical
- Instead, as time goes forward and with the benefit of tools, it is both essential and enough that we be SMEs

2

