



Dimensional Informatics

Tim Brailsford

The problem with data

A Relational Data Base

AUTHOR

au_id	au_lname	au_fname	address	city	state
172-32-1176	White	Johnson	10932 Bigge Rd.	Menlo Park	CA
213-46-8915	Green	Marjorie	309 63rd St. #411	Oakland	CA
238-95-7766	Carson	Cheryl	589 Darwin Ln.	Berkeley	CA
267-41-2394	O'Leary	Michael	22 Cleveland Av. #14	San Jose	CA
274-80-9391	Straight	Dean	5420 College Av.	Oakland	CA
341-22-1782	Smith	Meander	10 Mississippi Dr.	Lawrence	KS
409-56-7008	Bennet	Abraham	6223 Bateman St.	Berkeley	CA
427-17-2319	Dull	Ann	3410 Blonde St.	Palo Alto	CA
472-27-2349	Gringlesby	Burt	P0 Box 792	Covelo	CA
486-29-1786	Locksley	Charlene	18 Broadway Av.	San Francisco	CA

MENU

Graph Data System Quit
Window Status Page Hide

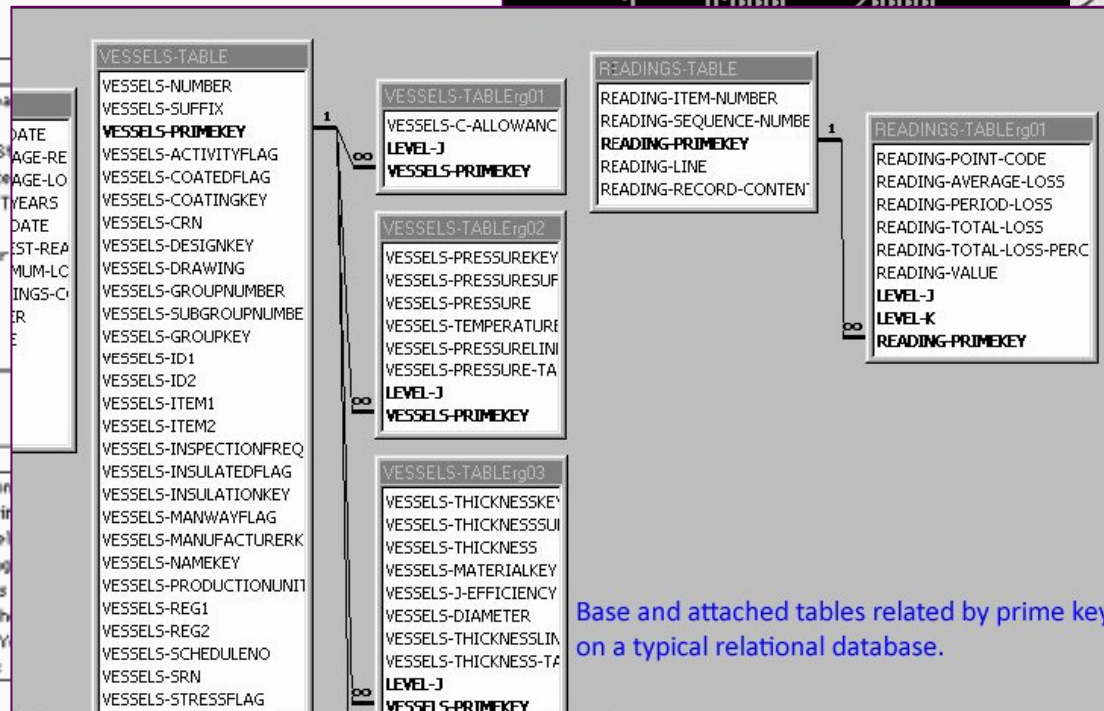
E	F	G
YEARS	SALARY	BONUS
2	40000	10000
3	45000	10000
4	75000	25000
3	65000	25000
5	65000	20000

TITLE

title_id	title
BU1032	The Busy Executive's Database
BU1111	Cooking with Computers
BU2075	You Can Combat Computer S
BU7832	Straight Talk About Comput
MC2222	Silicon Valley Gastronomic T
MC3021	The Gourmet Microwave
MC3026	The Psychology of Computer
PC1035	But Is It User Friendly?
PC8888	Secrets of Silicon Valley
PC9999	Net Etiquette
PS2091	Is Anger the Enemy?

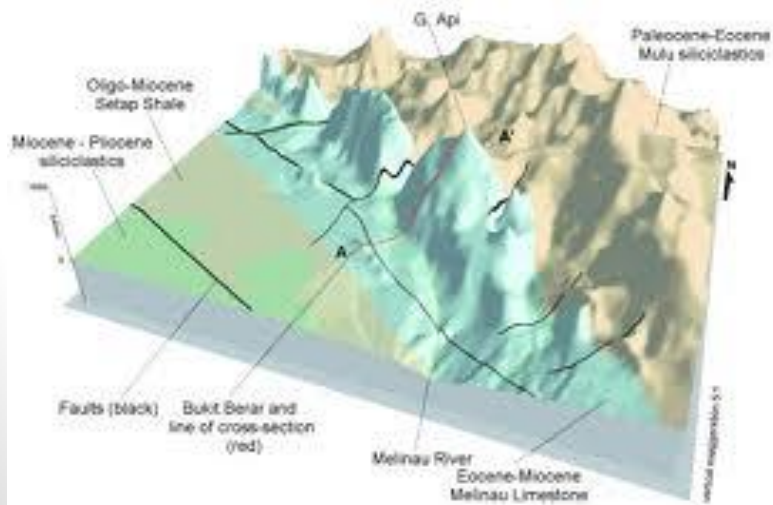
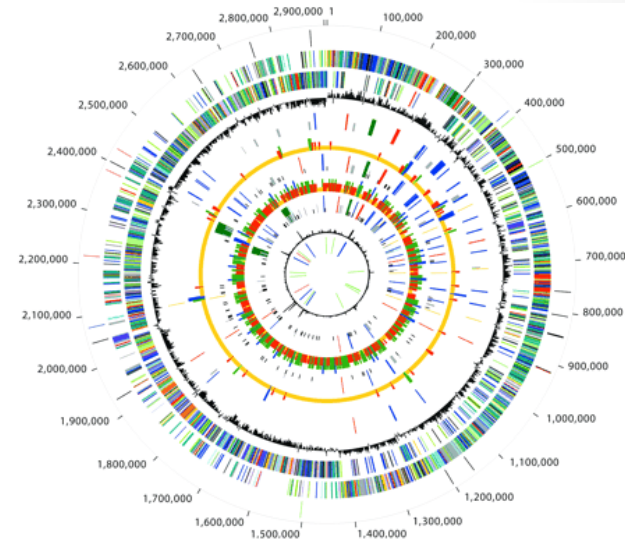
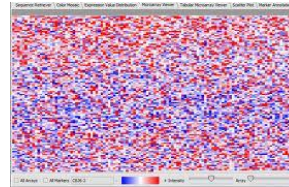
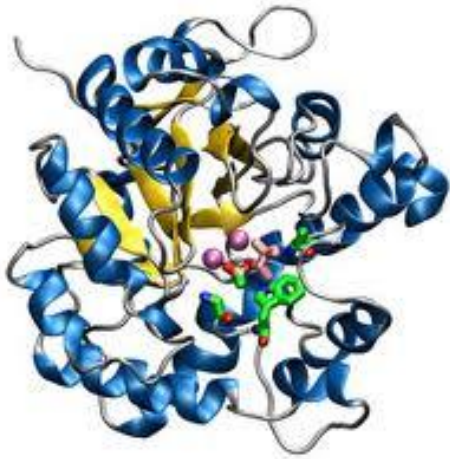
PUBLISHER

pub_id	pub_name	city
736	New Moon Books	Boston
977	Binnet & Hardley	Washing
1389	Algodata Infosystems	Berkeley
1622	Five Lakes Publishing	Chicago
1756	Ramona Publishers	Dallas
9901	GGG&G	Munich
9952	Scootney Books	New York
9999	Lucerne Publishing	Paris



Base and attached tables related by prime key on a typical relational database.

Why is this a problem?



The Null Conundrum

First Name	Middle Name	Last Name
Timothy	John	Brailsford
James	NULL	Goulding

Semi-Structured Data

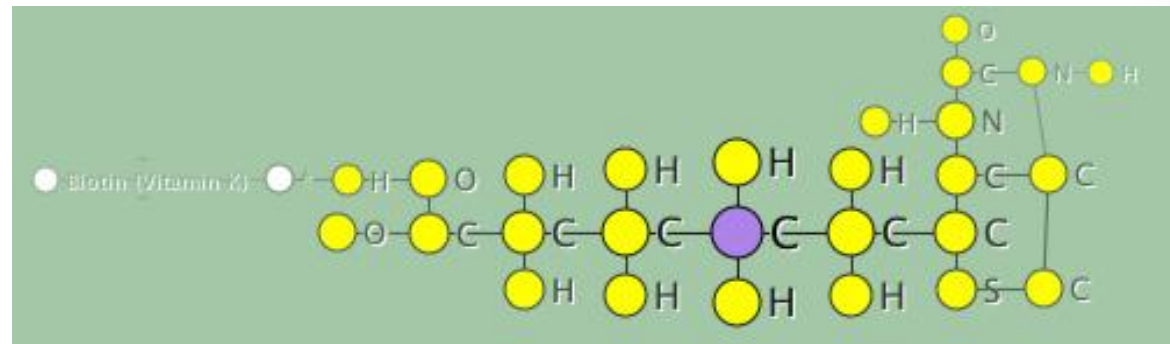
Dimensional Informatics

- Inspired by Ted Nelson's ZigZag System
- Data is contained in cells
- Cells are arranged in ordered lists – ranks in dimensions
- Dimensions usually have semantic meaning
- All cells exist in all dimensions – but the ordering can be different

Complex Data

H							He
Li	Be	B	C	N	O	F	Ne
Na	Mg	Al	Si	P	S	Cl	Ar
K	Ca	Ga	Ge	As	Se	Br	Kr

Element	Symbol	Atomic mass	Density	Melting poi
		g/mol	kg/m3	K
Hydrogen	H	1.01	0.09	14.0
Helium	He	4.00	0.17	1.0
Lithium	Li	6.94	534.00	452.0
Beryllium	Be	9.01	1800.00	1550.0
Boron	B	10.81	2500.00	2600.0
Carbon	C	12.01	2300.00	3800.0
Nitrogen	N	14.01	1.17	63.3
Oxygen	O	16.00	1.33	54.7
Fluorine	F	19.00	1.70	53.5
Neon	Ne	20.18	0.84	24.5
Sodium	Na	22.99	970.00	371.0
Magnesium	Mg	24.31	1741.00	924.0
Aluminium	Al	26.98	2700.00	933.2
Silicon	Si	28.09	2300.00	1680.0



Hyperorders

- James Goulding - PhD
- Any cell may be associated with any other in any dimension
- In set theory:

$$HM_{\text{hyperorder}} = \langle C, \{D_1, D_2, \dots, D_n\} \rangle$$

where $D_i \subset C_2$

- You then don't need a separate linking mechanism

Objectives

- CFFRC will have a lot of complex semi-structured data
- Develop a dimensional informatics storage layer - based on hyperorders for CFFRC datasets
- Implement this as a layer that can co-exist with conventional structures (i.e. RDMS)
- This should provide associations and inferences that would previously have been difficult or impossible

Supervision Team

- Tim Brailsford (UNMC)
- James Goulding (Horizon, UoN UK)
- Sean Mayes (CFFRC)

Approaches & Methodologies

- Continue the theoretical work
- Develop a DI server and associated query/retrieval language
- Test this with selected CFFRC data sets
- Implement an overlay to integrate the CFFRC knowledgebase with the DI server

Ideal student

- Really good programmer/software engineer
- Good knowledge of database theory
- Not afraid of maths!