

Allowing and Retrieving Constraint Violations in Data Storage

Lars Olson
DEG Lab
BYU Computer Science Dept.

The Problem

- * Compiling data for a single database
- * Some data may be contradictory or not make sense.
- * We may not be able to resolve the conflicts, so we want to store everything that *might* be true.

Current Databases

- * Can't store multiple values in one field without relaxing the constraint
- * Very few constraint types have been implemented
 - » Uniqueness
 - » Optional Fields
 - » Referential Integrity

Multiple Values

Multiple Birthdates

Joseph Hidson Harrison

b. 26 Jul 1782

26 Jan 1781

26 Jan 1782

Conflicting Name Information

James Hobday

married Sarah or Nancy Astley

Imprecise Values

Thomas Harrison

b. 22 Jan 1751 or ??? 1756

Thomas Harrison (father)

b. about 1705

James I of Scotland

b. Dec 1394

Single Value with Uncertainty

Sarah Green

b. about 6 Feb. 1743

Millicent Strelley

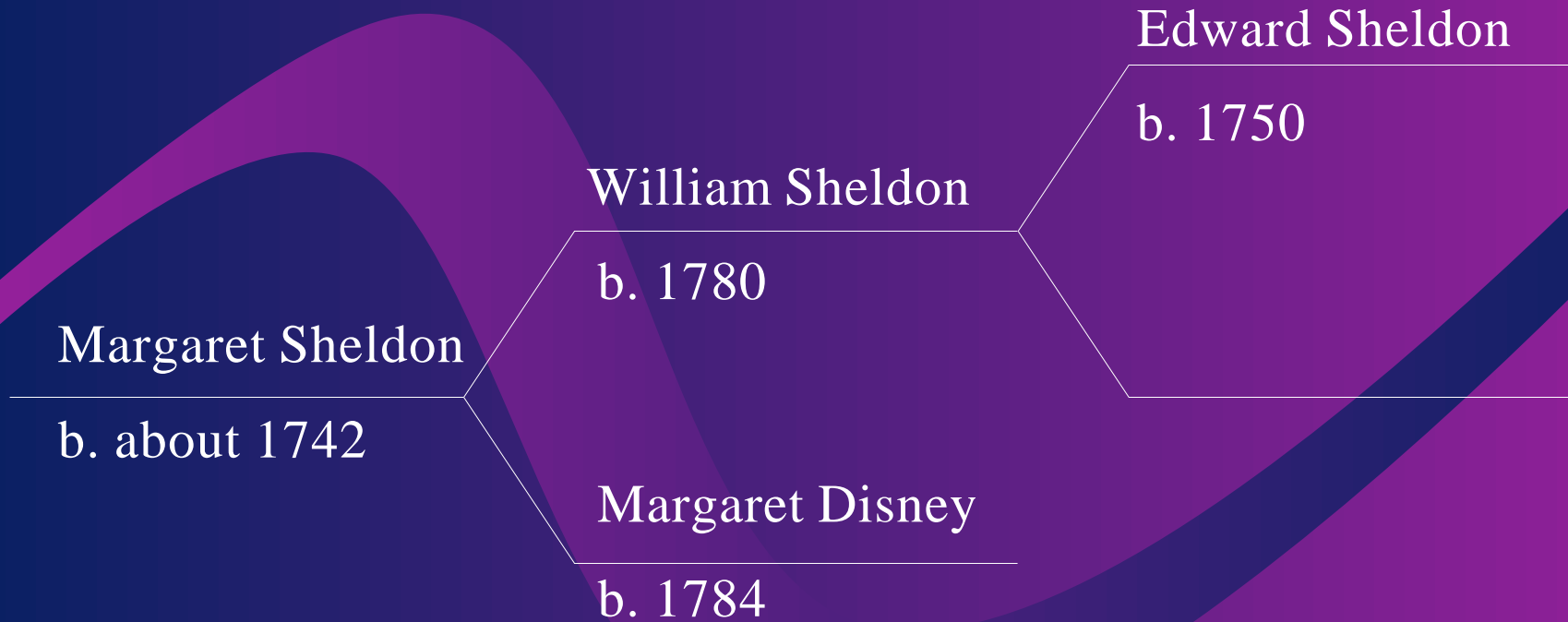
b. before 2 Jun. 1606

Henri Marchand

b. 1648

Caen?, Normandy?, France

Impossible Data



Cross-Products

James I of Scotland

b. Dec 1394

m. 2 Feb 1423 or 1424

d. 21 Feb 1436 or 1437

Subset of a Cross-Product

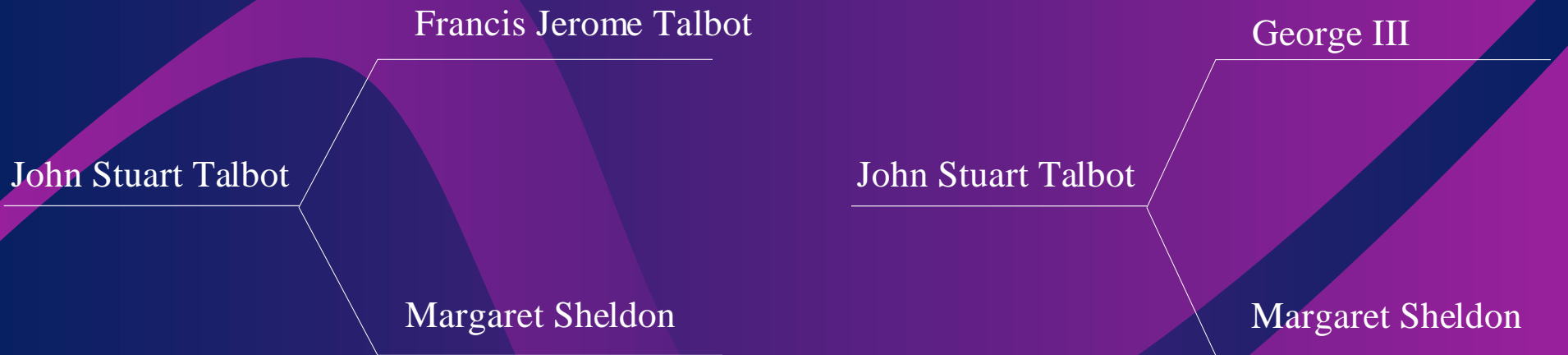
John Stuart Talbot's wife

Purcell Family
Westminster, England

< - ? - >

Loveridge Family
Oxford, England

Exceptions to the Model



perhaps we also need an optional explanation field in the model as well

Changing Information

Alexander Barron

b. 19 Jan 1779

Pendleton, 96th District, South Carolina

96th District is now called Anderson County

Inferred Information

William Bennett

b. ???

chr. 2 Jun 1703

Statistical Distributions

Olof Olsson

b. 1735

d. 24 Nov. 1822

87 years old?

Also:

- * Large families
- * Parents too old or too young to have children

Answering Queries

- * Find constraint violations
- * Answer normal queries in a way that "makes sense," whatever that means
 - » Get all ancestors born on June 17
 - Does that include those in June with an unknown day?
 - Does it include those that might have been born in June or in January?
 - What if we only know the year, or an approximate year?
 - » Get all ancestors born in June
 - What if we have a record with "June 17 or June 18"?
 - "June or January"? Year only?

What's Next

- * Previous Work (fuzzy databases, arbitration, ...)
- * Storing constraints and detecting violations
- * Implementation
- * Other applications -- military intelligence reports, weather data, astronomical data, forensics, ...