The Data Projection Model
Making Information Auditable

Michel Biezunski
Infoloom
(718) 921-0901
mb@infoloom.com
http://www.infoloom.com
About Michel Biezunski

- Created the Topic Maps paradigm and made it a standard (ISO 13250)
- Editor of the XTM specification. (2000-2001)
- Creator of the Data Projection Model.
- Major current projects and interest
  - IRS Tax Map
  - IEML (University of Ottawa)
  - Interest in bioinformatics and healthcare.
Introducing the Data Projection Model

- The Data Projection model enables information systems to become **auditable**. It aims at facilitating **system maintenance** and **knowledge management**. The Data Projection Model can also be used to **integrate** information assembled from a variety of sources and to **express multiple perspectives** on the same information set.

- This model represents a **synthesis between Topic Maps and RDF**, and can be **expressed in XML**. It will be presented **conceptually**, and the **application to TaxMap** will be discussed.
We can know:

- How much we are being paid and how much we pay for what we buy.
- How much money a public company is reporting.
- How the IRS calculates the amount of tax due.
- What is our credit rating.
- Values of stock, updated every day.
- The exchange rates with other currencies.
- etc.
Do we know:

- How does Google organize the search results?
- What software are used in voting machines?
- What the phone companies are doing with data they collect on our whereabouts?
- What processes are involved with dealing with identity theft?
- Where does a specific item of information come from?
- Who knows what?
Money vs. Information

● MONEY
  • Money flows through transactions
  • Money transactions are accounted for.
  • Accounting is about organizing the reporting.

● INFORMATION
  • Information flows through transactions.
  • Information transactions are not always accounted for.
  • There are no “accountants” for information content.
Bookkeeping

● Single Entry:
  • List of expenses per category
  • List of income per category.
  • Some money may be unaccounted for (although not desirable).

● Double Entry:
  • Organized by accounts.
  • Each transaction is an increase in one account matched with a decrease in another account. (Accounts are balanced.)
  • No money amount unaccounted for.
Double Entry Bookkeeping (1/4)

<table>
<thead>
<tr>
<th>Account Name</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets</td>
<td>Assets</td>
<td>$13,550.00</td>
</tr>
<tr>
<td>Current Assets</td>
<td>Current Assets</td>
<td>$13,550.00</td>
</tr>
<tr>
<td>Checking Account</td>
<td>Checking Account</td>
<td>$8,550.00</td>
</tr>
<tr>
<td>Savings Account</td>
<td>Savings Account</td>
<td>$5,000.00</td>
</tr>
<tr>
<td>Equity</td>
<td>Equity</td>
<td>$15,000.00</td>
</tr>
<tr>
<td>Expenses</td>
<td>Expenses</td>
<td>$1,450.00</td>
</tr>
<tr>
<td>Travel and Entertainment</td>
<td>Travel and Entertainment</td>
<td>$1,450.00</td>
</tr>
<tr>
<td>Travel</td>
<td>Travel</td>
<td>$1,450.00</td>
</tr>
<tr>
<td>Air</td>
<td>Plane Tickets (Business Trips)</td>
<td>$450.00</td>
</tr>
<tr>
<td>Hotel</td>
<td>Hotel (Business Trips)</td>
<td>$1,000.00</td>
</tr>
<tr>
<td>Income</td>
<td>Income</td>
<td>$0.00</td>
</tr>
<tr>
<td>Liabilities</td>
<td>Liabilities</td>
<td>$0.00</td>
</tr>
</tbody>
</table>

$, Grand Total: Assets: $13,550.00  Profits: -$1,450.00
Double Entry Bookkeeping (2/4)
Double Entry Bookkeeping (3/4)
Modeling Techniques

- There are several ways to model this.
- Binary relations is one among them:
• Entity-Relationship Model

● Can always be decomposed into binary relations.

DPM & XML

● Binary Relations:
  • Decompose into pieces / Recompose into views

● SGML & XML
  • Separate presentation from content. Apply styles afterwards.

<table>
<thead>
<tr>
<th>Debit</th>
<th>Description</th>
<th>Credit</th>
<th>Category</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Equity: Opening Balance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Savings Account</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Checking Account</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hotel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trip to San Francisco</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expenses: Travel:Air</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expenses: Travel:Hotel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expenses: Travel:Hotel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expenses: Travel:Hotel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>450$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From Multiple to Multiple Via One

- **SGML / XML**
  - One source,
  - Multiple outputs

  \((\text{Ex Uno Plures})\)

- **DPM**
  - Diverse inputs,
  - One common representation,
  - Multiple outputs

  \((\text{E Pluribus Plures Via Unum})\)
What the Data Projection Model does

- Solve Integration Problems Between Various Classification Systems.
- Flexible Network instead of Rigid Hierarchies
- Auditing Information Networks
- Enabling Multiple Perspectives
- Bottom-Up Applications
- Maintaining Complex, Multidimensional Information Models
How does the Data Projection Model work?

- Captures Semantic Relations.
- Captures Processes.
- Networks Information Components.
- Enables Maintenance and Navigation.
Flattening the World
Perspectives are used in projections:

- Different ways to go from 3D to 2D.
- Different points of view.

Description: World in Mercator projection, Source: Kober-Kümmerly+Frey Media AG
Real World Information:

- Is multidimensional.
- Can be flattened to be processed.
- **Binary relations** correspond to **2D space**
- Translating a world of n-ary relations into a world of binary relations is a kind of *projection*.
- **Perspective** is what accompanies projection from n-ary relations to binary relations.
A “perspector” is noted:

\[< x \mid o \mid y >\]

\(x\) and \(y\) are operands (order matters).

\(o\) is an operator.

A perspector can represent anything, for example:

\[< \text{New York} \mid \text{is a} \mid \text{city} >\]

(This is an instance/class relationship)

\[< \text{city} \mid \text{added to the system by} \mid \text{MB} >\]

(This is usually considered metadata), etc.
Perspector Notation: Example

<Equity:Opening Balance| Debit | T2 >
< Savings Account | Credit | T2 >
<T2 | Amount | 5000$>
<Equity:Opening Balance | Debit | T1 >
< Checking Account | Credit | T1>
<T1 | Amount | 10000$>

<Checking Account | Debit | T3 >
<Checking Account | Debit | T4>
<Expenses:Travel:Hotel| Credit| T3>
<Expenses:Travel:Air|Credit|T4>
<T3 Amount|1000$>
<T4 Amount|450$>
<T3 Description|Hotel>
<T4 Description|Trip to San Francisco>
<Expenses:Travel:Hotel|Category|Expenses etc.

Debit  Description
Credit Category
Amount
Auditability: 2 + 3 not 5

\[
< 2 \ | \ + \ | \ 3 >
\]

is the addition of 2 and 3.

- We are interested not by the result, but by the fact that the two numbers, 2 and 3, are being combined together through the operator “+”.
- Recording this information enables us to trace back the origin of any item. Here we will know where 5 comes from.
DPM and RDF

- RDF is based on triples that express statements: subject – object – predicate
- RDF connects URLs
- RDF statements are not automatically reified.

- DPM is based on triples that express operations: x operand – operator – y operand
- DPM is not limited to URLs
- DPM perspectors are automatically reified.
DPM and Topic Maps

- Topic Maps is a Navigation system using topics as nodes for representing subjects.
- Names, Types, Occurrences are topics connected through specific relationships.

- DPM is a Navigation system based on nodes
- All nodes are related with other nodes.
- Topic Maps can be considered an application of DPM.
DPM and XML

< Washington | is a | city >

can be written:

<perspector id="p1">
<x>Washington</x><o>is a</o><y>city</y>
</perspector>

This is XML.
But there is more.
Example: Name versus Subject

A Name does not identify a Subject:

• Variant names may be used to designate the same subject.
  • Synonyms
  • Typographical variations
• One name may identify several subjects.
Names

- Wash D.C.
- Washington, DC
- General Washington
- Denzel Washington
- George Washington
- Wa
- Washington State

Washington
## Names

<table>
<thead>
<tr>
<th>Washington</th>
<th>is an alternate name for</th>
<th>Wash. D.C. &gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washington</td>
<td>is an alternate name for</td>
<td>Washington, DC &gt;</td>
</tr>
<tr>
<td>Washington</td>
<td>is an alternate name for</td>
<td>General Washington &gt;</td>
</tr>
<tr>
<td>Washington</td>
<td>is an alternate name for</td>
<td>George Washington &gt;</td>
</tr>
<tr>
<td>Washington</td>
<td>is an alternate name for</td>
<td>Wa &gt;</td>
</tr>
<tr>
<td>Washington</td>
<td>is an alternate name for</td>
<td>Washington State &gt;</td>
</tr>
<tr>
<td>Washington</td>
<td>is an alternate name for</td>
<td>Denzel Washington &gt;</td>
</tr>
</tbody>
</table>
Emerging Subjects

- Wash D.C.
- Washington, DC
- General Washington
- George Washington
- Washington
- Wa
- Denzel Washington
- Washington State
- Washington State
Strings Become Subjects

- Wash D.C.
- Washington, DC
- General Washington
- George Washington
- Washington
- Wa
- Washington State
- Denzel Washington
- Washington State
Generalization

Wash D.C. is a name for Washington, DC
General Washington is a name for Washington
George Washington is a name for Washington State
Denzel Washington is a name for Washington
Names and Subjects

< Washington    | is a name for     | _city_of_Washington >
< Washington DC | is a name for     | _city_of_Washington >
< Wash. D.C.    | is a name for     | _city_of_Washington >
< Washington    | is a name for     | _General_G_Washington >
< General Washington | is a name for  | _General_G_Washington >
< George Washington | is a name for | _General_G_Washington >
< Washington    | is a name for     | _Washington_State >
< Wa             | is a name for     | _Washington_State >
< Washington State | is a name for     | _Washington_State >
< Washington    | is a name for     | _Denzel_Washington >
< Denzel Washington | is a name for | _Denzel_Washington >
Strings as Subjects

< Washington    | is in character set | UTF-8        >
< Washington    | is a name for      | _city_of_Washington >
< Washington    | is a name in the language | English >
Integration

- abbreviates
- also known as
- represents
- is a name for
- is a code name for
- is usually called
- stands for
- designates
- indicates
- is the last name of

- Wash D.C.
- Washington, DC
- General Washington
- Washington
- George Washington
- Wash State
- Denzel Washington
- Washington State
Diversity

< _city_of_Washington
< Washington DC
< Wash. D.C.
< Washington
< General_G_Washington
< George Washington
< Washington
< Wa
< Washington State
< Washington
< Denzel Washington

is usually called
indicates
abbreviates
is a name for
also_known_as
represents
stands for
is a code name for
is a name for
is last name of
designates

Washington >
_city_of_Washington >
_General_G_Washington
_General_Washington >
_Washington_State >
_Washington_State >
_Denzel_Washington >
_Denzel_Washington >
## Perspective on Naming

| `<city_of_Washington>` | `<Washington DC>` | is a name for | `<city_of_Washington>` |
| `<Wash. D.C.>` | is a name for | `<city_of_Washington>` |
| `<Washington>` | is a name for | `<city_of_Washington>` |
| `<General_G_Washington>` | is named | `<General_G_Washington>` |
| `<George_Washington>` | is a name for | `<General_G_Washington>` |
| `<Washington>` | is a name for | `<General_G_Washington>` |
| `<Wa>` | is a name for | `<Washington_State>` |
| `<Washington State>` | is a name for | `<Washington_State>` |
| `<Washington>` | is a name for | `<Washington_State>` |
| `<Denzel_Washington>` | is a name for | `<Denzel_Washington>` |
Multidimensional Information

< New York | is a name for _New_York_City >
< New York | is a name for _New_York_State >
< New York | is a name for _New_York_County >
< New York | is a name for _Manhattan >
< New York | is a name for _Wall_Street >
< New York | is an old name for _Manhattan >
< Nueva York | is a name for _New_York_City >
< ניו יורק | is a name for _New_York_City >
< New York | is a name in the language _English >
< Nueva York | is a name in the language _Spanish >
< New York | is a name in the language _French >
< English | is a name for _English >
< Anglais | is a name in the language _French >
< Anglais | is a name in the language _English >
< Inglés | is a name for _English >
< Inglés | is a name in the language _Spanish >

etc., etc., etc., etc., etc., etc., etc., etc., etc., etc., etc., etc.
An Example of Auditing using the Data Projection Model

- TaxMap is a Topic Map application developed for the IRS since 2001 to help taxpayer assistors navigate publications, forms and instructions in terms of the subjects with which they are concerned.
Operations on Names

- TaxMap is built by a combination of automatic and manual processes. Names are added, modified, sometimes deleted, or regarded as synonyms.
- It's hard to know where a topic name comes from.
### Income Earned Abroad

**Index**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>created</td>
<td>from document p593-001</td>
</tr>
<tr>
<td>2</td>
<td>related containment</td>
<td>Income</td>
</tr>
<tr>
<td>3</td>
<td>related PFT</td>
<td>Living Abroad</td>
</tr>
<tr>
<td>4</td>
<td>normalizes as</td>
<td>INCOME EARNED ABROAD</td>
</tr>
<tr>
<td>5</td>
<td>URL</td>
<td>ts0/tp6717.htm</td>
</tr>
</tbody>
</table>
U.S. Citizens and Resident Aliens Abroad

American citizens abroad
Living Abroad

Forms and Instructions

**Form 2350** Application for Extension of Time to File U.S. Income Tax Return

**Form 2555** Foreign Earned Income

  2555 Instructions

Publications

**Publication 54** Tax Guide for U.S. Citizens and Resident Aliens Abroad

**Publication 516** U.S. Government Civilian Employees Stationed Abroad

**Publication 593** Tax Highlights for U.S. Citizens and Residents Going Abroad

Links Inside Publications

**Publication 17** - Your Federal Income Tax - Filing Information

  U.S. Citizens and Residents Living Outside the United States ...
Where does “Living Abroad” come from?

<table>
<thead>
<tr>
<th>Index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>alternate name PFT</td>
</tr>
<tr>
<td>2</td>
<td>American citizens abroad</td>
</tr>
<tr>
<td>3</td>
<td>alternate name PFT</td>
</tr>
<tr>
<td>4</td>
<td>U.S. Citizens and Resident Aliens Abroad</td>
</tr>
<tr>
<td>5</td>
<td>privileged name is</td>
</tr>
<tr>
<td>6</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>7</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>8</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>9</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>10</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>11</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>12</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>13</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>14</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>15</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>16</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>17</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>18</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>19</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>20</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>21</td>
<td>replaced PFT by</td>
</tr>
<tr>
<td>22</td>
<td>physical presence by</td>
</tr>
<tr>
<td>23</td>
<td>LIVING ABROAD</td>
</tr>
<tr>
<td>24</td>
<td>Key Topic</td>
</tr>
</tbody>
</table>
## Containment Rule Results

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>862</td>
<td>Roth IRAs and traditional IRAs.</td>
</tr>
<tr>
<td>863</td>
<td>Roth IRAs and traditional IRAs.</td>
</tr>
<tr>
<td>864</td>
<td>Discount on Debt Instruments</td>
</tr>
<tr>
<td>865</td>
<td>Hope and Lifetime Learning Credits</td>
</tr>
<tr>
<td>866</td>
<td>Health Spa Expenses</td>
</tr>
<tr>
<td>867</td>
<td>Health Spa Expenses</td>
</tr>
<tr>
<td>868</td>
<td>Exceptions to reporting OID.</td>
</tr>
<tr>
<td>869</td>
<td>Exceptions to reporting OID.</td>
</tr>
<tr>
<td>870</td>
<td>SSN on correspondence.</td>
</tr>
<tr>
<td>871</td>
<td>Capital Expense</td>
</tr>
<tr>
<td>872</td>
<td>Home Mortgage</td>
</tr>
<tr>
<td>873</td>
<td>Home Mortgage</td>
</tr>
<tr>
<td>874</td>
<td>Maximum Exclusion</td>
</tr>
<tr>
<td>875</td>
<td>Income from property given to a child.</td>
</tr>
<tr>
<td>876</td>
<td>Income from property given to a child.</td>
</tr>
<tr>
<td>877</td>
<td>Income from property given to a child.</td>
</tr>
<tr>
<td>878</td>
<td>Medicare Advantage MSAs</td>
</tr>
<tr>
<td>879</td>
<td>Farm Income, cash method</td>
</tr>
<tr>
<td>880</td>
<td>Farm Income, cash method</td>
</tr>
<tr>
<td>881</td>
<td>Farm Income, cash method</td>
</tr>
<tr>
<td>882</td>
<td>Farm Income, cash method</td>
</tr>
</tbody>
</table>

If one topic name is entirely contained into another one, they get automatically related.
Synonyms Created by Tax Experts

139 < Gambling
140 < Gift, Property
141 < Gift, Shares
142 < Gifts to Charity
143 < HCTC
144 < Help from IRS
145 < Hiring Employees
146 < Hobby Expenses
147 < Home Mortgage Interest
148 < Home Mortgage Interest
149 < Home Office
150 < Home Vacation
151 < Hope Credit
152 < Household Employment
153 < IC-DISC
154 < IRA
155 < IRA
156 < ITIN
157 < Important Reminder
158 < Important Reminder
159 < Important Reminder
160 < Important Reminder

Wagering
Property Received as a Gift
Shares Acquired by Gift
Charitable Contributions
Health Coverage Tax Credit
Taxpayer Assistance
Hiring New Employees
Activity not for profit
Itemized Deduction, Home Mortgage Interest
Mortgage Interest, Home
Business Use of Home
Vacation Home
Credit, Hope
Employment, Household
Interest Charge Domestic International Sales Corporation
Individual Retirement Account
Individual Retirement Arrangement
Taxpayer Identification Number
Reminders, Important
Reminders, Important
Reminders, Important
More Information

Demos, other presentations available at:
http://www.infoloom.com

Michel Biezunski
Infoloom

(718) 921-0901
mb@infoloom.com