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

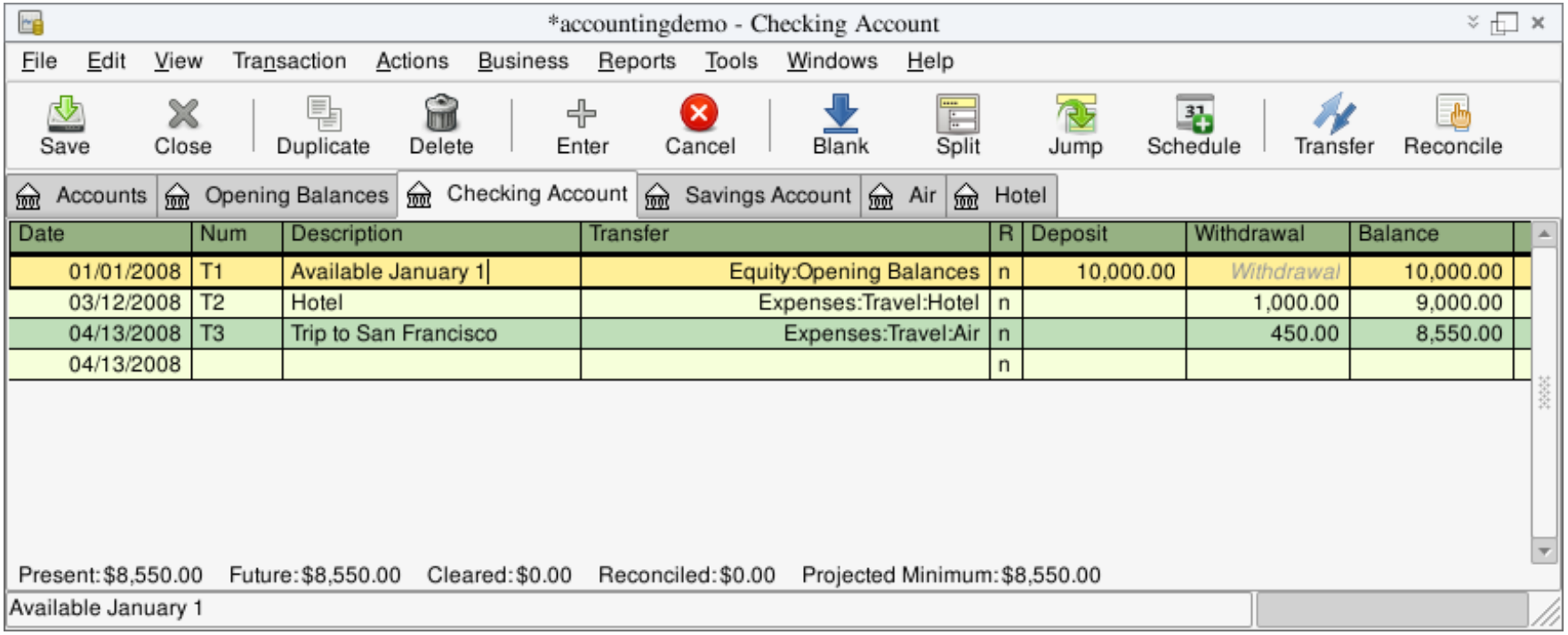
- MONEY
  - Money flows through transactions
  - Money transactions are accounted for.
  - Accounting is about organizing the reporting.
- INFORMATION
  - Information flows through transactions.
  - Information transactions are not always accounted for.
  - There are no “accountants” for information content.

Bookkeeping

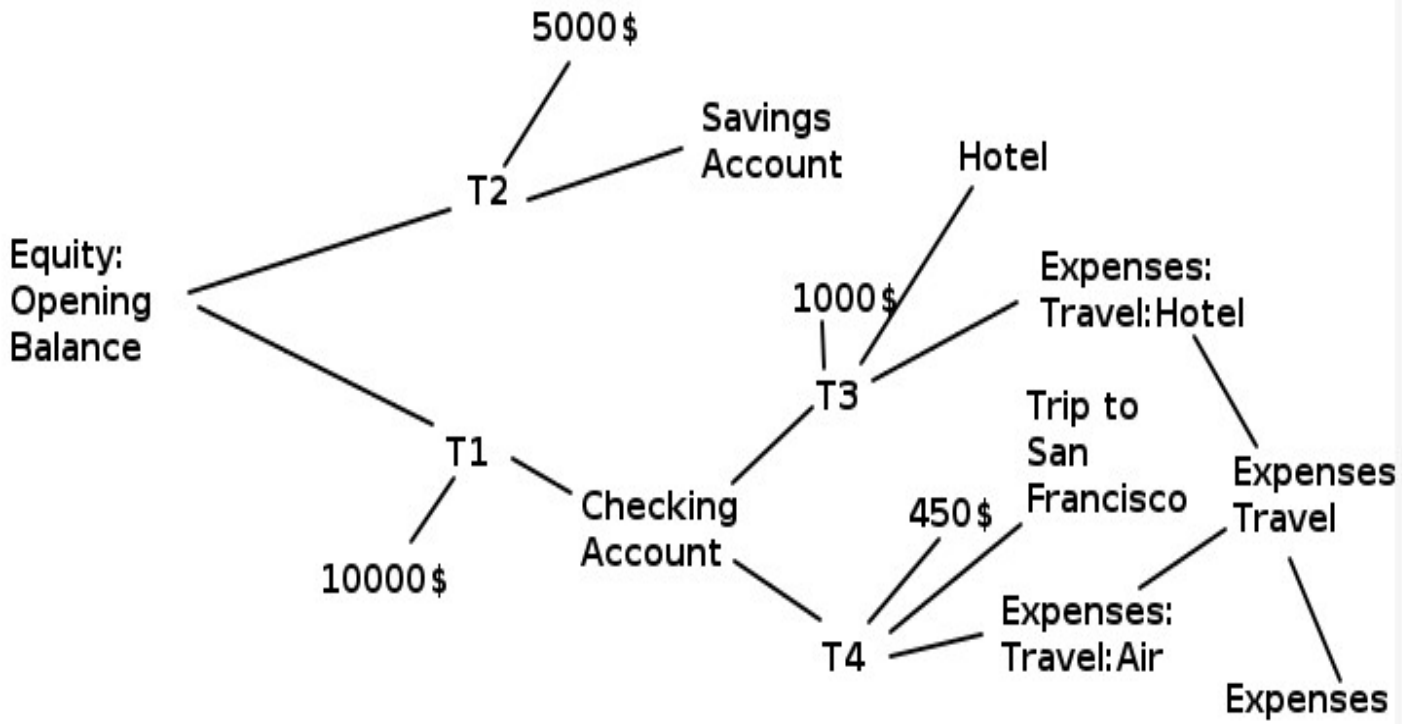
- Single Entry:
  - List of expenses per category
  - List of income per category.
  - Some money may be unaccounted for (although not desirable).
- Double Entry:
  - Organized by accounts.
  - Each transaction is an increase in one account matched with a decrease in another account. (Accounts are balanced.)
  - No money amount unaccounted for.

Double Entry Bookkeeping (1/4)

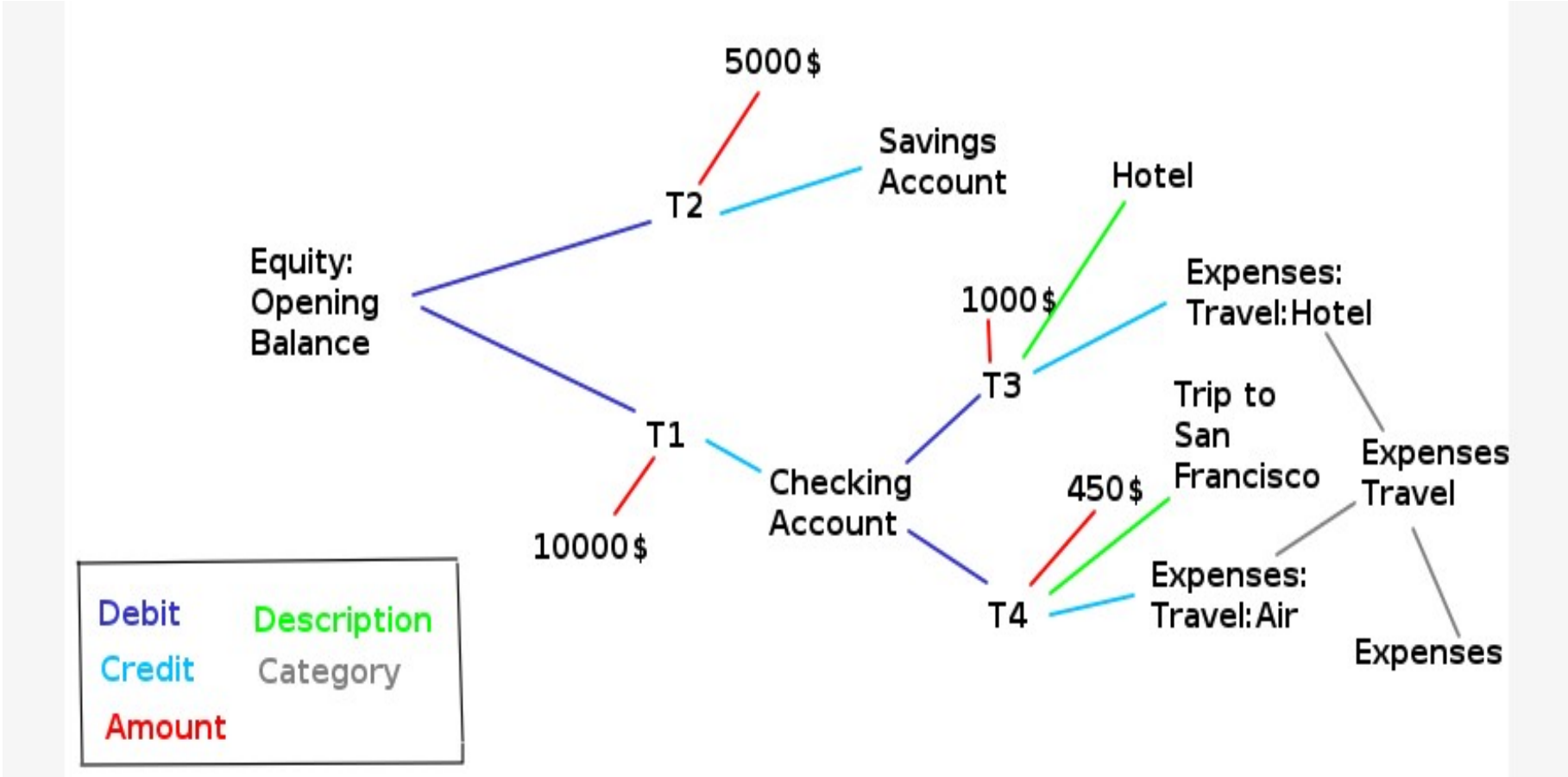
Double Entry Bookkeeping (2/4)



Double Entry Bookkeeping (3/4)

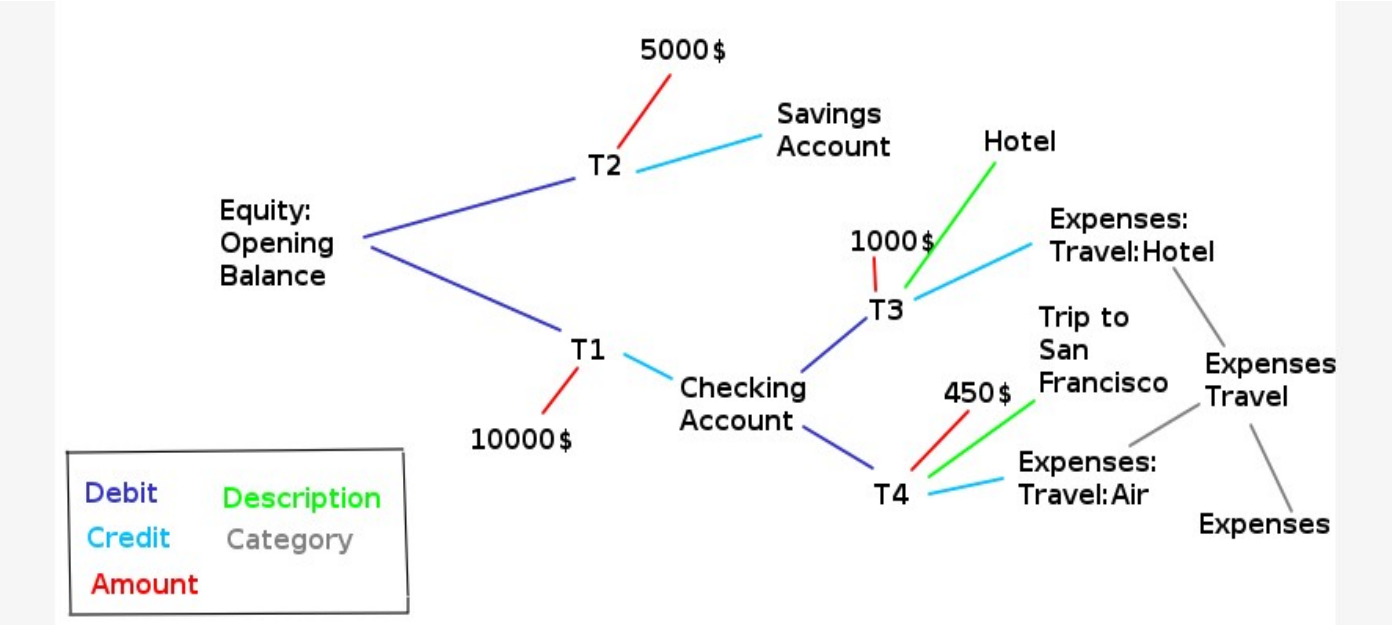


Double Entry Bookkeeping (4/4)



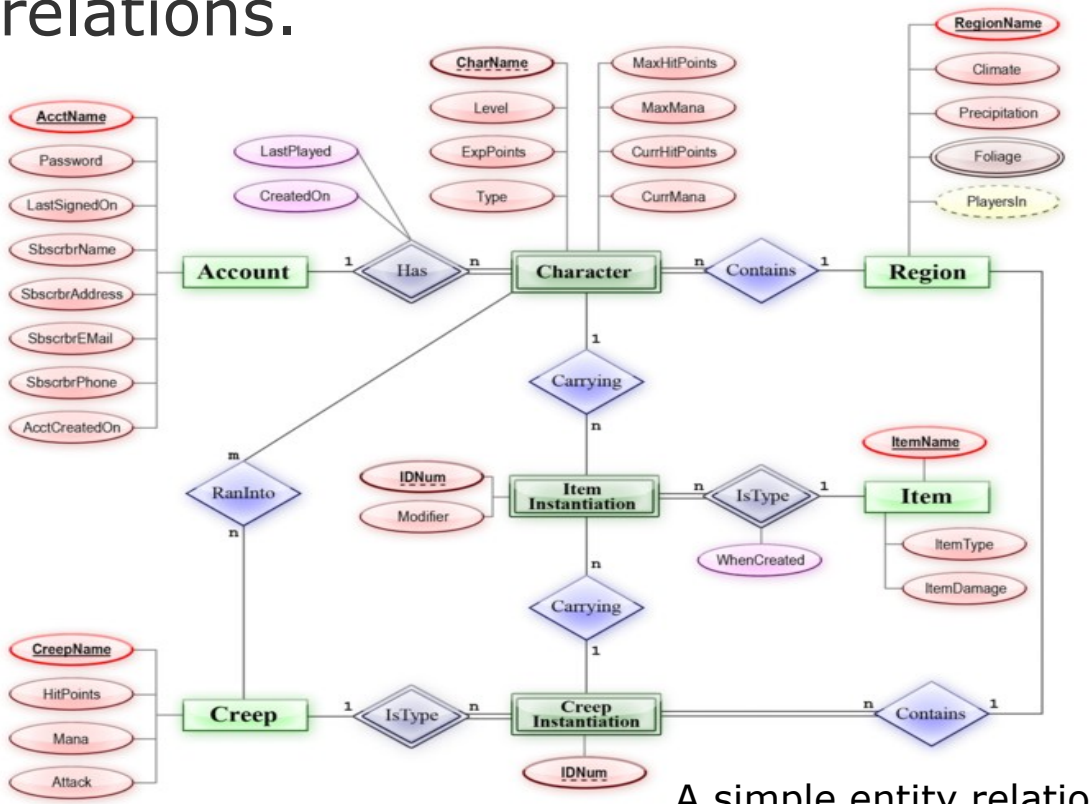
Modeling Techniques

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



Entity-Relationship Model

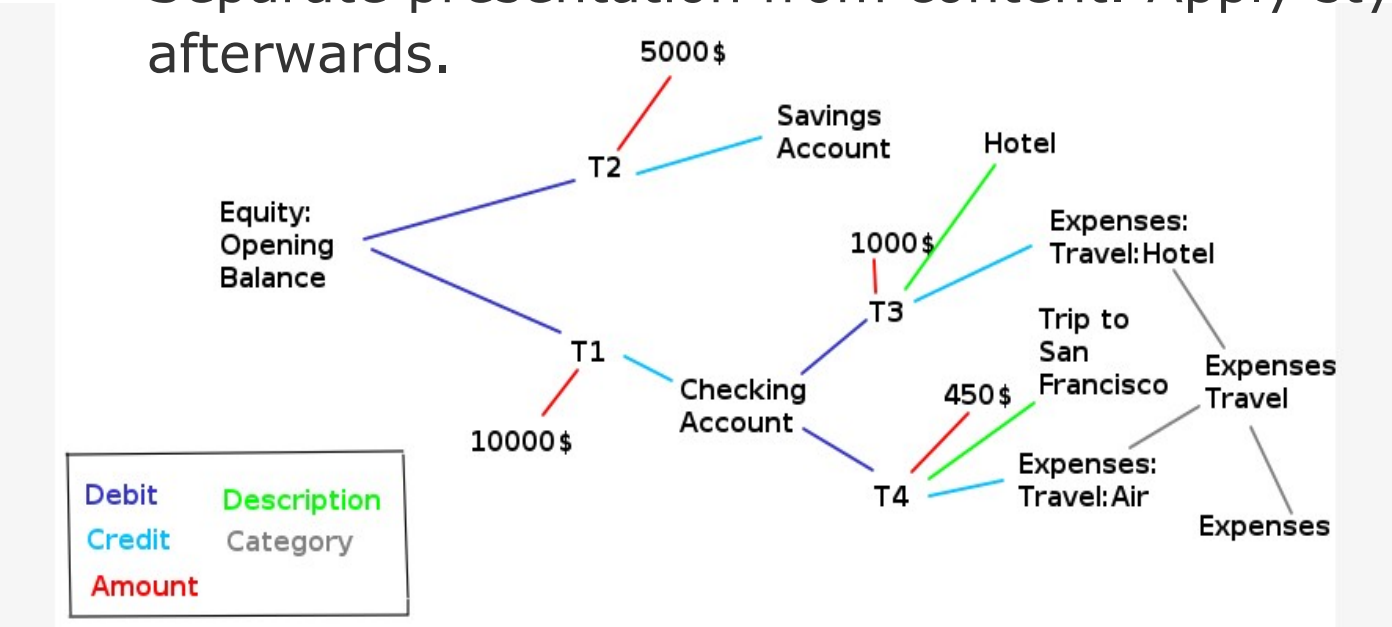
- Can always be **decomposed** into binary relations.



A simple entity relationship model.
http://en.wikipedia.org/wiki/Entity-relationship_model

DPM & XML

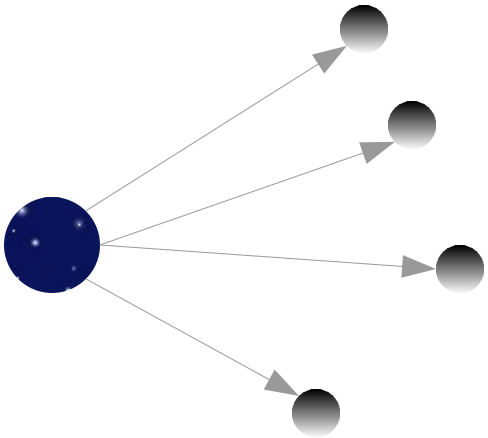
- Binary Relations:
  - Decompose into pieces / Recompose into views
- SGML & XML
  - Separate presentation from content. Apply styles afterwards.



From Multiple to Multiple Via One

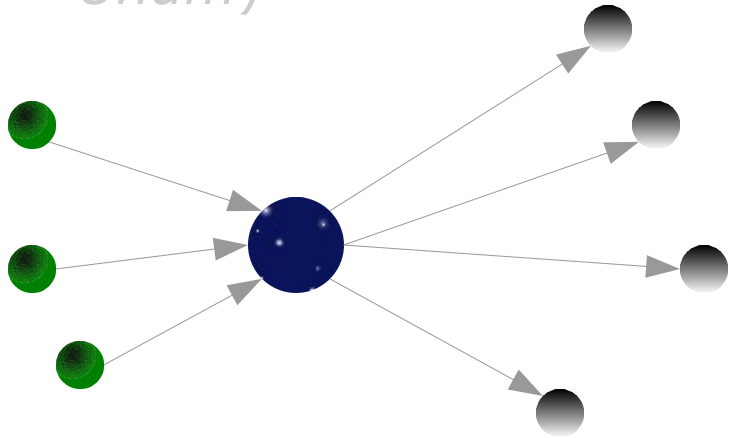
- SGML / XML
  - One source,
  - Multiple outputs

  (*Ex Uno Plures*)



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

  (*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

Projection

- 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 Date: 21.11.2005, http://en.wikipedia.org/wiki/Image:Welt_Mercator_Atlantik.png

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.

Perspector

A "**perspector**" is noted:

`< x | o | y >`

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

`o` is an operator.

A perspector can represent anything, for example:

`< New York | is a | city >`

(This is an instance/class relationship)

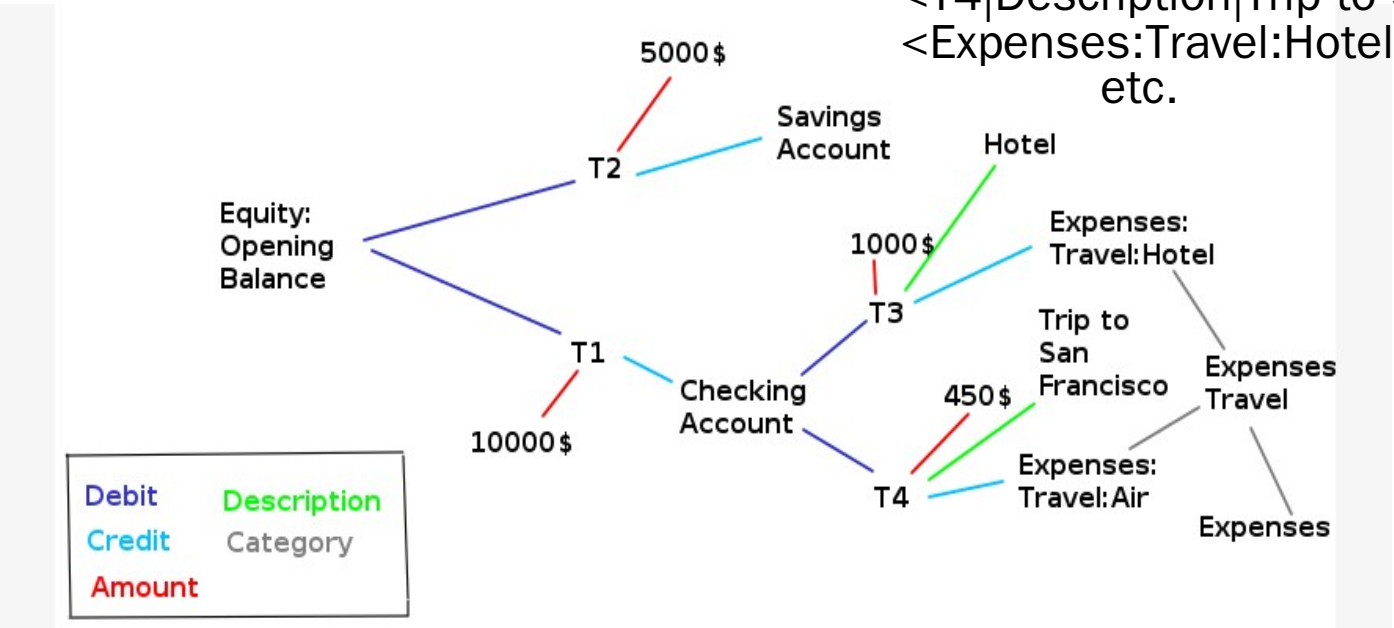
or `< city | added to the system by | MB >`

(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 Travel>
etc.



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.

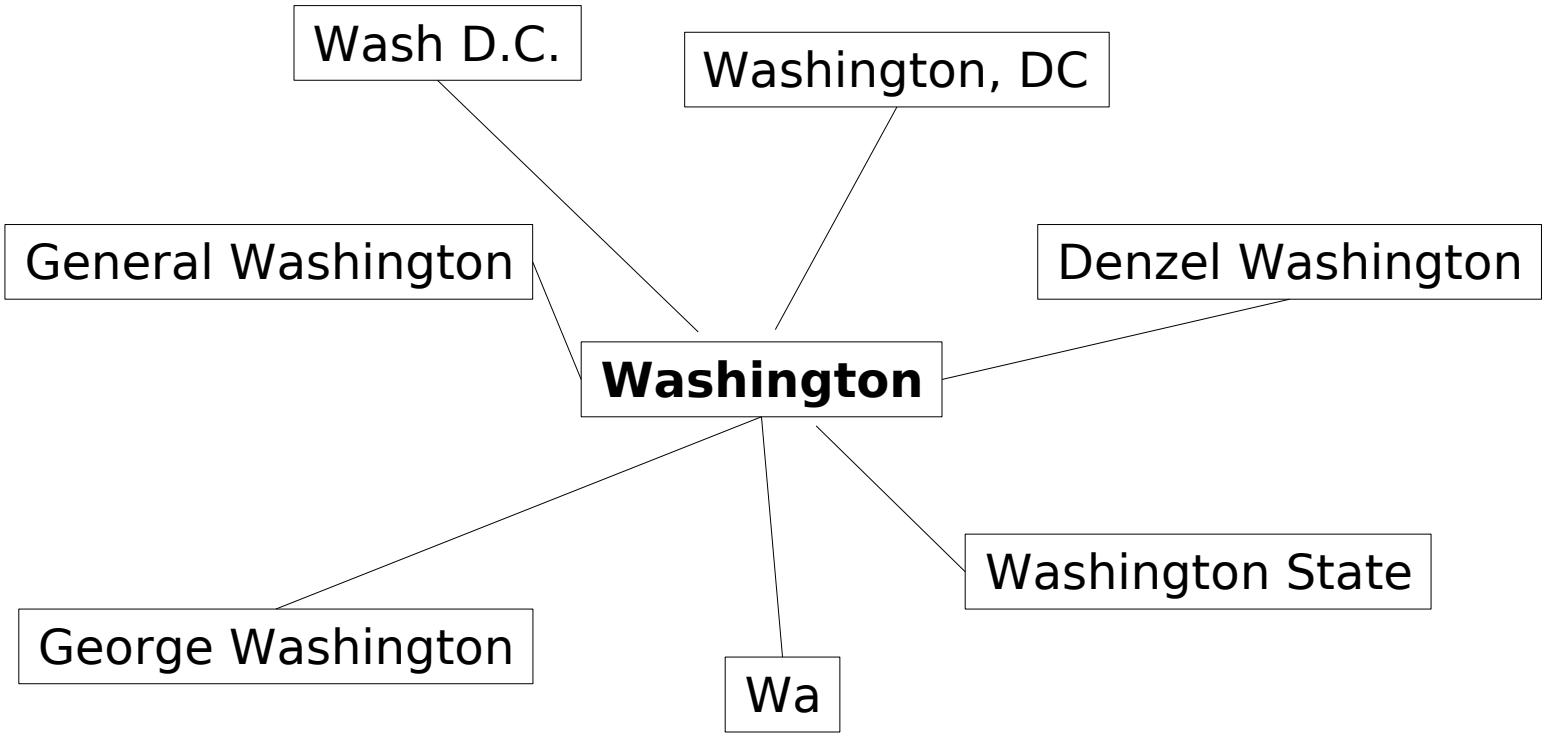
But there is more.

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.

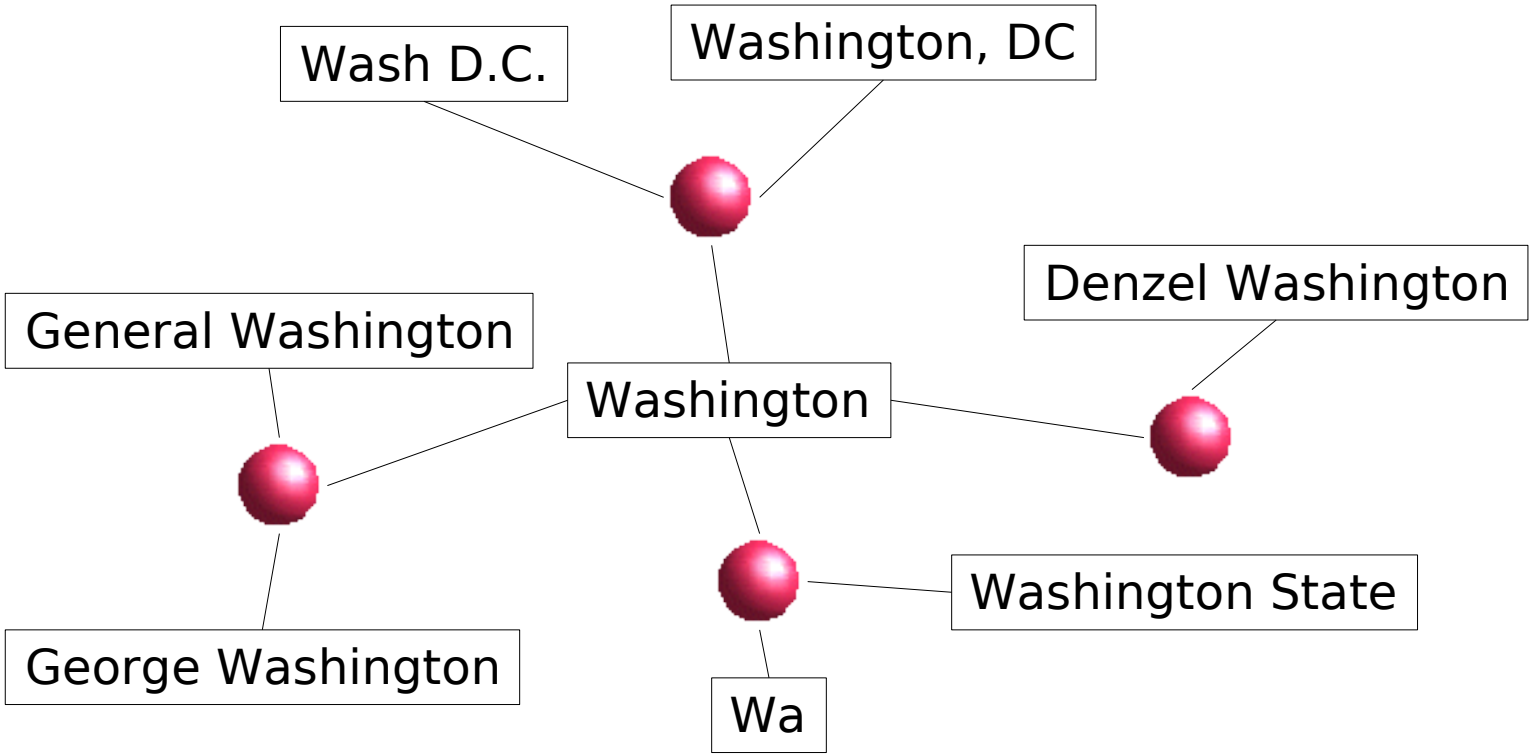
Names


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

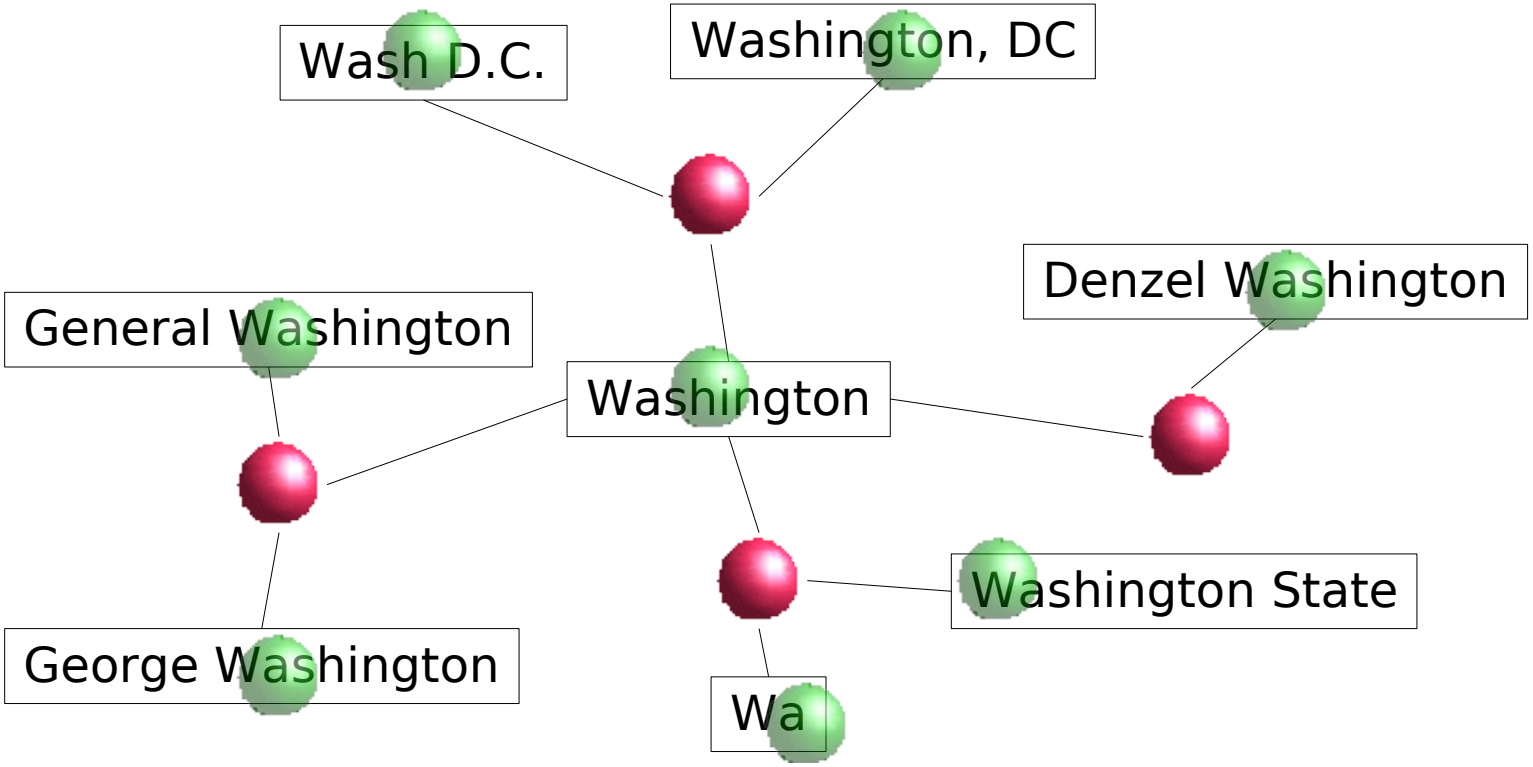
Names

```
< Washington | is an alternate name for | Wash. D.C. >
< Washington | is an alternate name for | Washington, DC >
< Washington | is an alternate name for | General Washington>
< Washington | is an alternate name for | George Washington >
< Washington | is an alternate name for | Wa >
< Washington | is an alternate name for | Washington State >
< Washington | is an alternate name for | Denzel Washington >
```

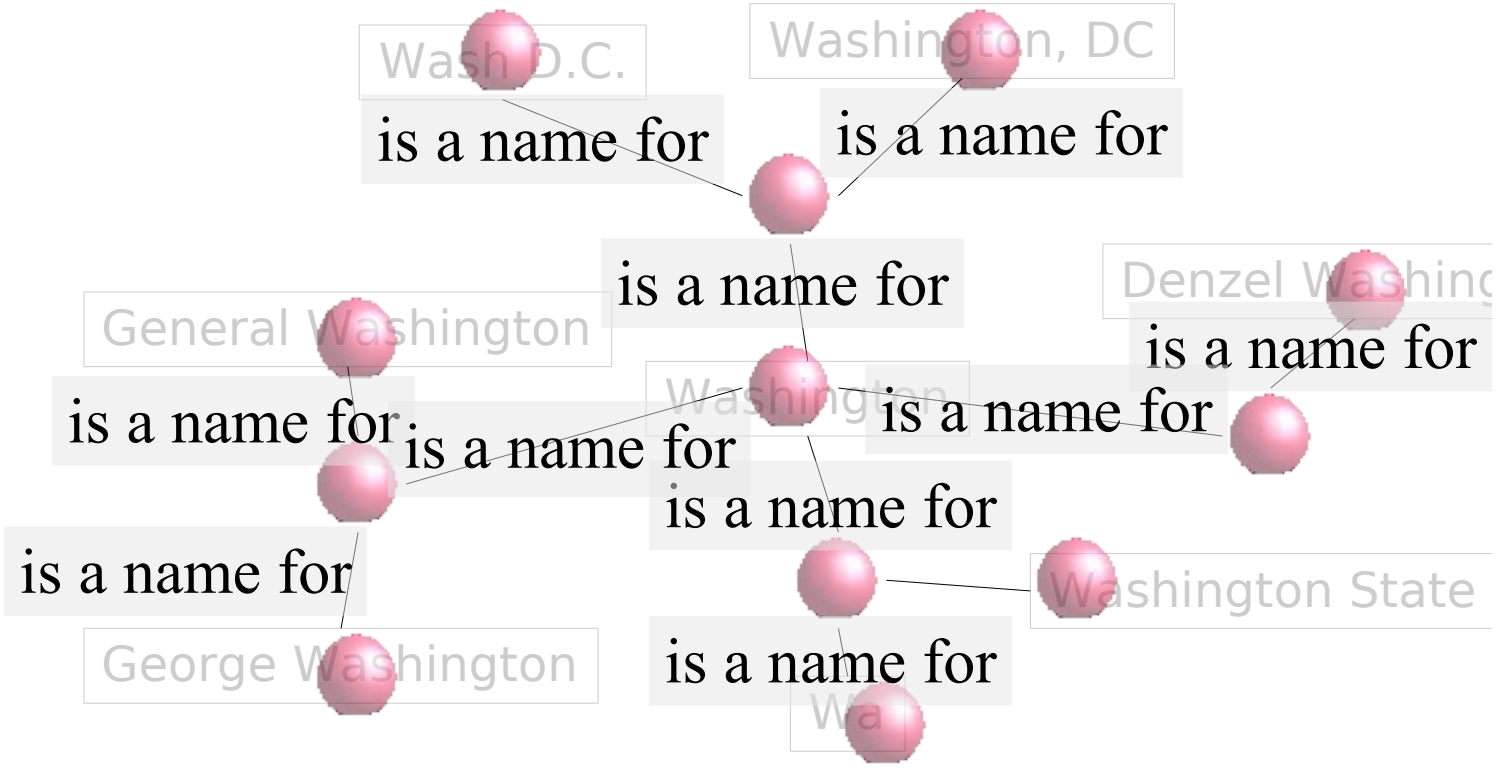
Emerging Subjects



Strings Become Subjects



Generalization



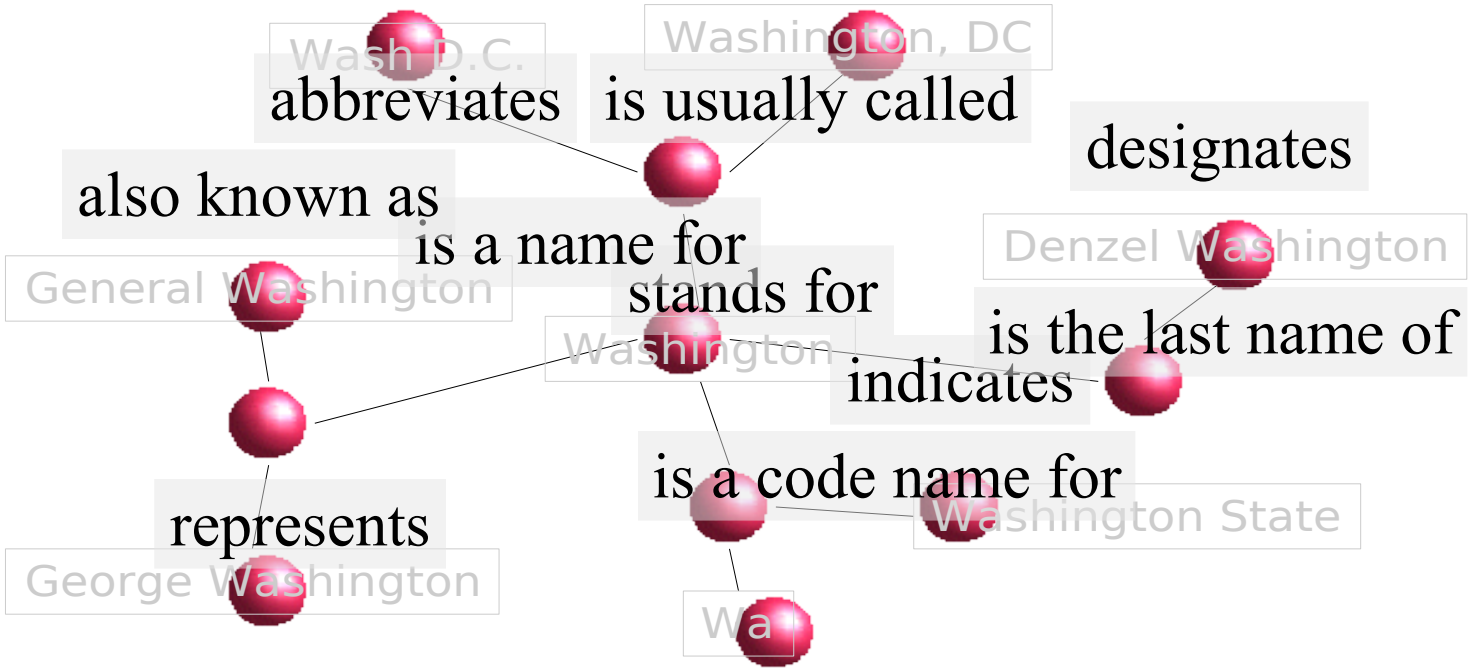
Names and Subjects

```
< 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



Diversity

< _city_of_Washington	is usually called	Washington >
< Washington DC	indicates	_city_of_Washington >
< Wash. D.C.	abbreviates	_city_of_Washington >
< Washington	is a name for	_General_G_Washington
<_General_G_Washington	also_known_as	General Washington >
< George Washington	represents	_General_G_Washington
< Washington	stands for	_Washington_State >
< Wa	is a code name for	_Washington_State >
< Washington State	is a name for	_Washington_State >
< Washington	is last name of	_Denzel_Washington >
< Denzel Washington	designates	_Denzel_Washington >

Perspective on Naming

< _city_of_Washington	is named	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_G_Washington	is named	General 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 >

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 >
< נו 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 >
< English	is a name in the language	_English >
< Anglais	is a name for	_English >
< Anglais	is a name in the language	_French >
< 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., etc.

An Example of Auditing using the Data Projection Model

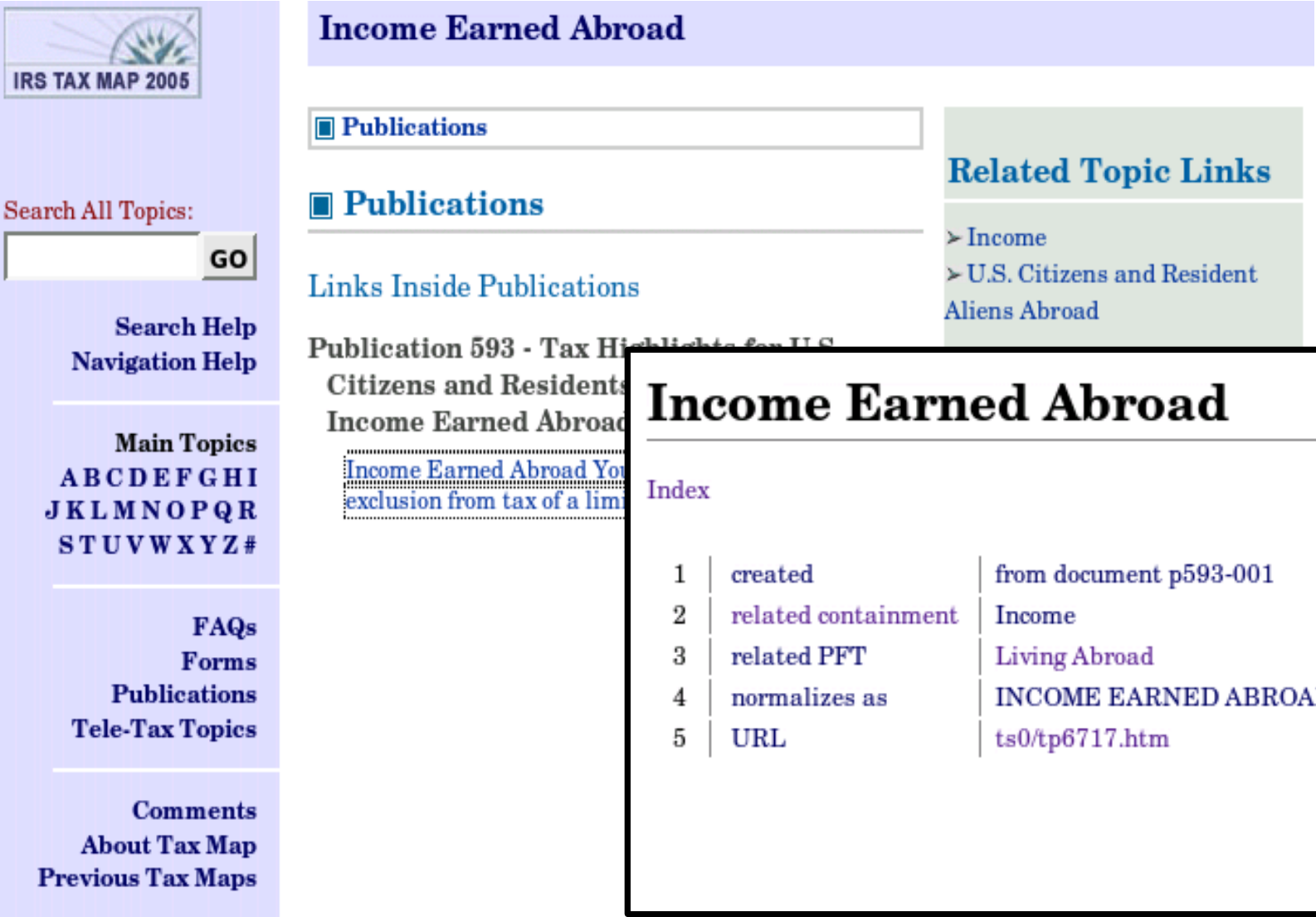
- TaxMap is a Topic Map application developed for the IRS since 2001 to help taxpayer assistors navigate publications, forms and instructions in terms of the subjects with which they are concerned.

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.

Tax Map Audited:

Income Earned Abroad



U.S. Citizens and Resident Aliens Abroad

American citizens abroad
Living Abroad

Search All Topics:
GO

Search Help
Navigation Help

Main Topics
A B C D E F G H I
J K L M N O P Q R
S T U V W X Y Z #

FAQs
Forms
Publications
Tele-Tax Topics

Comments
About Tax Map
Previous Tax Maps

▣ Forms ▣ Publications ▣ FAQs ▣ TeleTax

▣ Forms and Instructions

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

Form 2555 Foreign Earned Income MORE

2555 Instructions

▣ Publications

Related Publications

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

Publication 516 U.S. Government Civilian Employees Stationed Abroad PDF MORE

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

Links Inside Publications

Publication 17 - Your Federal Income Tax - Filing Information

U.S. Citizens and Residents Living Outside the United States ...

Related Topic Links

- Alien, Resident
- Bona Fide Residence Test
- Employment Abroad
- Foreign Corporation
- Foreign Income
- Income Earned Abroad
- International Issues
- Physical Presence Test
- U.S. Citizen
- U.S. Citizens and Resident and Nonresident Aliens
- VISA Status

Where does "Living Abroad" come from?

Living Abroad

Index

1		alternate name PFT	American citizens abroad	>
2	< U.S. Citizens and Resident Aliens Abroad	alternate name PFT		
3		privileged name is	U.S. Citizens and Resident Aliens Abroad	>
4	< Individuals Abroad	replaced PFT by		
5	< U.S. Citizen Abroad	replaced PFT by		
6	< U.S. citizens abroad	replaced PFT by		
7	< U.S. Citizens and Residents Living Outside the United States	replaced PFT by		
8	< U.S. Citizens or Residents Living Abroad	replaced PFT by		
9		related PFT	Bona Fide Residence Test	>
10	< Employment Abroad	related PFT		
11		related PFT	F2350	>
12		related PFT	Foreign Corporation	>
13		related PFT	Foreign Income	>
14	< Income Earned Abroad	related PFT		
15	< P516	related PFT		
16	< P54	related PFT		
17	< P593	related PFT		
18		related PFT	Physical Presence Test	>
19	< Retirees who were working abroad.	related PFT		
20	< Survivors of decendents who were working abroad.	related PFT		
21	< VISA Status	related PFT		
22		normalizes as	LIVING ABROAD	>
23		URL	ts0/tp16663.htm	>
24		topic type PFT	Key Topic	>

Containment Rule Results

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

If one topic nam
is entirely
contained
into another one
they get
automatically
related.

Synonyms Created by Tax Experts

139 < Gambling	Wagering	>
140 < Gift, Property	Property Received as a Gift	>
141 < Gift, Shares	Shares Acquired by Gift	>
142 < Gifts to Charity	Charitable Contributions	>
143 < HCTC	Health Coverage Tax Credit	>
144 < Help from IRS	Taxpayer Assistance	>
145 < Hiring Employees	Hiring New Employees	>
146 < Hobby Expenses	Activity not for profit	>
147 < Home Mortgage Interest	Itemized Deduction, Home Mortgage Interest	>
148 < Home Mortgage Interest	Mortgage Interest, Home	>
149 < Home Office	Business Use of Home	>
150 < Home Vacation	Vacation Home	>
151 < Hope Credit	Credit, Hope	>
152 < Household Employment	Employment, Household	>
153 < IC-DISC	Interest Charge Domestic International Sales Corporation	>
154 < IRA	Individual Retirement Account	>
155 < IRA	Individual Retirement Arrangement	>
156 < ITIN	Taxpayer Identification Number	>
157 < Important Reminder	Reminders, Important	>

More Information

Demos, other presentations available at:

http://www.infoloom.com

Michel Biezunski

Infoloom

(718) 921-0901

mb@infoloom.com