

The Data Projection Model

Making Information Auditable

Michel Biezunski

Infoloom

(718) 921-0901
mb@infoloom.com

<http://www.infoloom.com>

Contents

The Data Projection Model

- What it's for.

- What it is.

- Where it comes from.

- How to use it.

Why Bother?

	Agree?	
● Mess is a fact of life. We can't get rid of it.	Yes	No
● Universal agreement? Forget it!	Yes	No
● Freedom of speech is here to stay.	Yes	No
● Computers don't really understand what we want, no matter what.	Yes	No
● We are not sure that we are finding what we need.	Yes	No
● Transparency is good.	Yes	No
● Privacy should be preserved.	Yes	No

What the Data Projection Model is for

- 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

What the Data Projection Model does

- Captures Semantic Relations.
- Captures Processes.
- Networks Information Components.
- Enables Maintenance and Navigation.

A Flat World



Perspective

- Art: methods to represent 3-dimensional space on a flat surface.
- Geometry: laws of perspective express what is invariant according to various points of view.

Projection



- Perspectives are used in projections:
 - Different ways to go from 3D to 2D.
 - Different points of view.
- Once projected, the world is flat.

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:

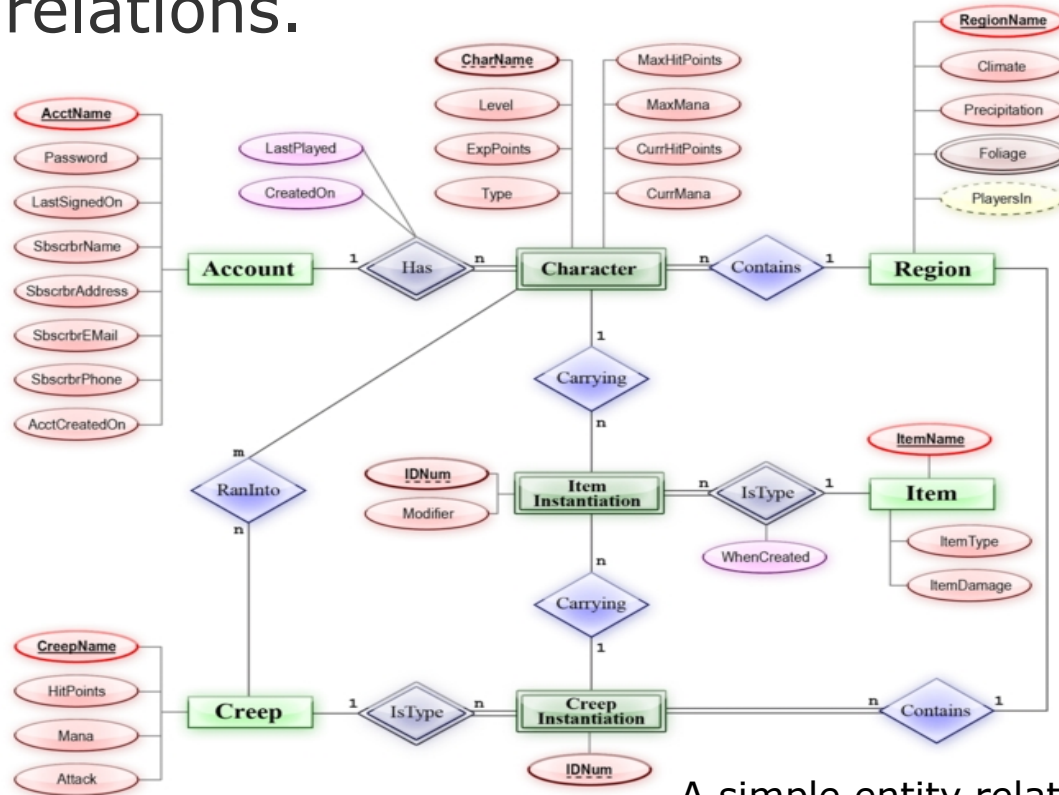
- Is multidimensional.
- Is flattened to be processed.
- There are multiple ways to flatten information.
- There are multiple ways to look at information after it has been flattened.
- We are interested by knowing which one is being used in the system we are using.

A Flat Information World

- 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.

•Multidimensional Information

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



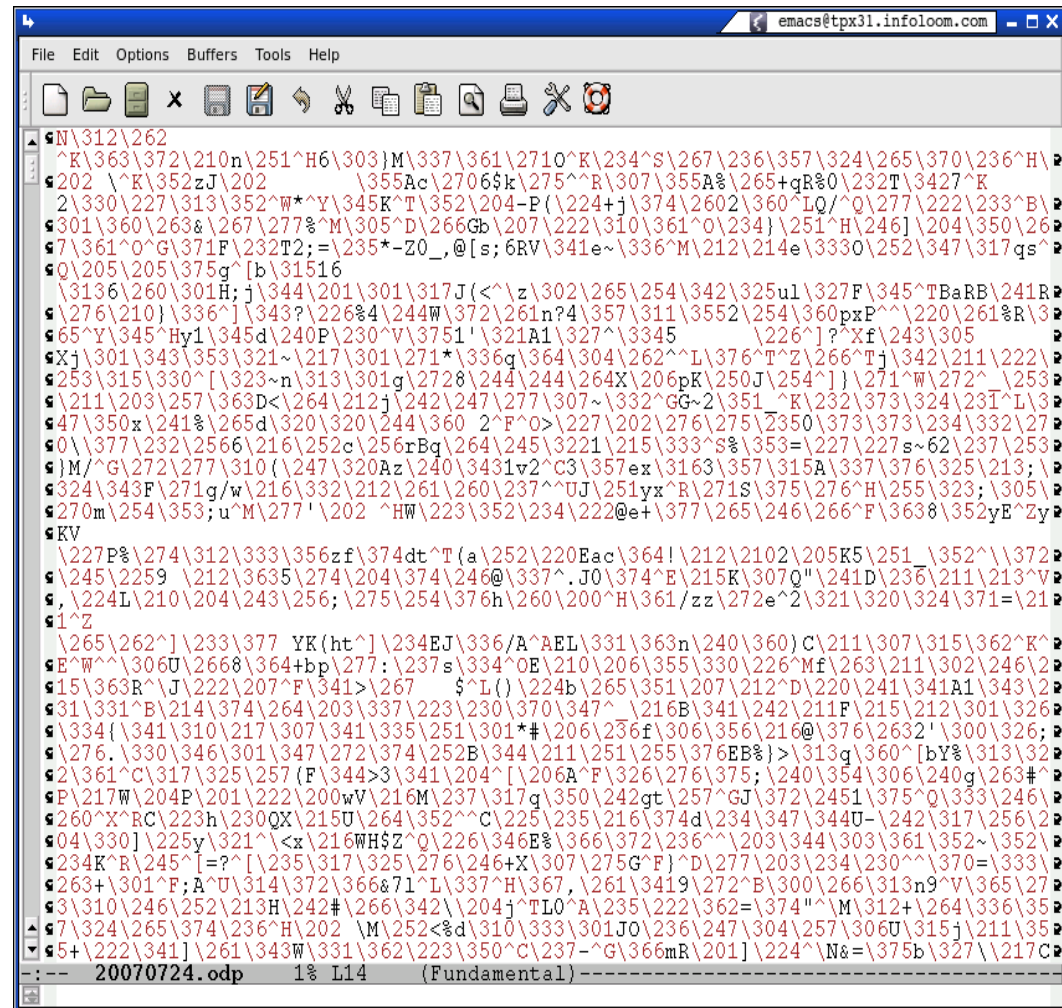
A simple entity relationship model.
http://en.wikipedia.org/wiki/Entity-relationship_model

Equivalents in Other Fields

- Computer Science
- Chemistry
- Accounting

Computer Science

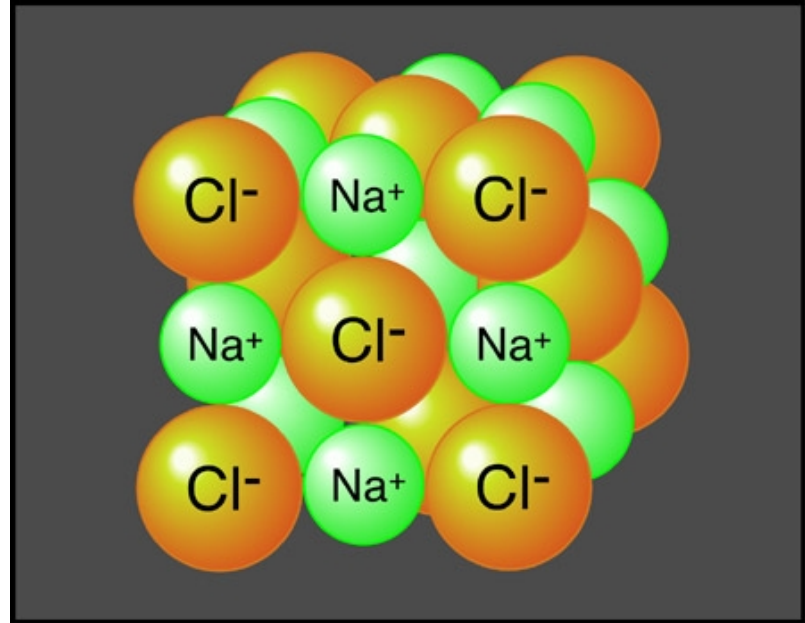
- High level Languages
- User Interfaces
- Assembly Language:
 - 0s and 1s
- Internal Formats:
 - 0s and 1s



```
File Edit Options Buffers Tools Help
N\312\262
^K\363\372\210n\251^H6\303}M\337\361\2710^K\234^S\267\236\357\324\265\370\236^H\
202 ^\K\352zJ\202 \355Ac\2706$K\275^R\307\355A% \265+qR%0\232T\3427^K
2\330\227\313\352^W^Y\345K^T\352\204-P(\224+j\374\2602\360^LQ/^Q\277\222\233^B\
301\360\263&\267\277%M\305^D\266Gb\207\222\310\361^O\234}\251^H\246}\204\350\26
7\361^O^G\371F\232T2;=\235*-Z0_,@[s;6RV\341e~\336^M\212\214e\3330\252\347\317qs^
Q\205\205\375g^[b\31516
\3136\260\301H;j\344\201\301\317J(<^z\302\265\254\342\325ul\327F\345^TBaRB\241R
\276\210}\336^j\343?\226%4\244W\372\261n?4\357\311\3552\254\360pxP^^\220\261%R\3
65^Y\345^Hy1\345d\240P\230^V\3751'\321A1\327^ \3345 \226^]?^Xf\243\305
Xj\301\343\353\321~\217\301\271*\336q\364\304\262^L\376^T^Z\266^Tj\342\211\222\
253\315\330^[ \323~n\313\301g\2728\244\244\264X\206pK\250J\254^j)\271^w\272^ \253
211\203\257\363D<\264\212j\242\247\277\307~\332^GG~2\351 ^K\232\373\324\231^L\3
47\350x\241%\265d\320\320\244\360 2^F^O>\227\202\276\275\2350\373\373\234\332\27
0\377\232\2566\216\252c\256rBq\264\245\3221\215\333^S%\353=\227\227s~62\237\253
M/^G\272\277\310(\247\320Az\240\3431v2^C3\357ex\3163\357\315A\337\376\325\213;\
324\343F\271g/w\216\332\212\261\260\237^UJ\251yx^R\271S\375\276^H\255\323;\305\
270m\254\353;u^M\277'\202 ^HW\223\352\234\222@e+\377\265\246\266^F\3638\352yE^Zy
KV
\227P%\274\312\333\356zf\374dt^T(a\252\220Eac\364!\212\2102\205K5\251 \352^ \372
\245\2259 \212\3635\274\204\374\246@\337^J0\374^E\215K\307Q"\241D\236\211\213^V
\242L\210\204\243\256;\275\254\376h\260\200^H\361\zz\272e^2\321\320\324\371=\21
1^Z
\265\262^]\233\377 YK(ht^]\234EJ\336/A^AEL\331\363n\240\360)C\211\307\315\362^K^
E^W^~\306U\2668\364+bp\277:\237s\334^OE\210\206\355\330\226^mf\263\211\302\246\2
45\363R^J\222\207^F\341>\267 $^L() \224b\265\351\207\212^D\220\241\341A\343\2
31\331^B\214\374\264\203\337\223\230\370\347^ \216B\341\242\211F\215\212\301\326
334\341\310\217\307\341\335\251\301*#\206\236f\306\356\216@\376\2632'\300\326;
276 \330\346\301\347\272\374\252B\344\211\251\255\376EB%}>\313q\360^ [bY% \313\32
2\361^C\317\325\257(F\344>3\341\204^[\206A^F\326\276\375;\240\354\306\240g\263#^
P\217W\204P\201\222\200wV\216M\237\317q\350\242gt\257^GJ\372\2451\375^Q\333\246\
260^X^RC\223h\230QX\215U\264\352^C\225\235\216\374d\234\347\344U~\242\317\256\2
04\330]\225y\321^<x\216WH$Z^Q\226\346E% \366\372\236^ \203\344\303\361\352~\352\
234K^R\245^ [= ?^ [\235\317\325\276\246+X\307\275G^F]^D\277\203\234\230^ \370=\333\
263+\301^F;A^U\314\372\366&71^L\337^H\367 \261\3419\272^B\300\266\313n9^V\365\27
3\310\246\252\213H\242#\266\342\204j^TL0^A\235\222\362=\374^"M\312+\264\336\35
7\324\265\374\236^H\202 \M\252<%d\310\333\301J0\236\247\304\257\306U\315\211\35
5+\222\341\261\343W\331\362\223\350^C\237~^G\366mR\201\224^N&=\375b\327\217C
-- 20070724.odp 1% L14 (Fundamental) --
```

Chemistry

- Matter decomposed into atoms.
- Atoms composed into molecules.



Atomic representation of sodium chloride or table salt.
Source: <http://www.physicalgeography.net/>. Quoted in
Michael Pidwirny, <http://www.eoearth.org/article/Matter>

Accounting

- Double Entry Accounting
 - Record = Transaction Between Accounts
 - Checks and Balances

 Accounts		 Online Services					
Date	Num	Description	Transfer	R	Expense	Rebate	Balance
07/23/2007 ▾	Num	Domain Name Renewal	Assets:Checking Account	n	232.00	Rebate	232.00
07/23/2007				n			

Binary Relations

A “perspector”

$$< x \mid o \mid y >$$

can represent information semantics:

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

or can represent a process:

$$< \text{city} \mid \text{added in the system by} \mid \text{MB} >$$

x and y are operands: order matters.

o is an operator.

2 + 3 not 5

- $< 2 \mid + \mid 3 >$
- 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 “Plus”.
- Recording this information enables us to trace back the origin of any item. Here we will know why 5 is what it is.

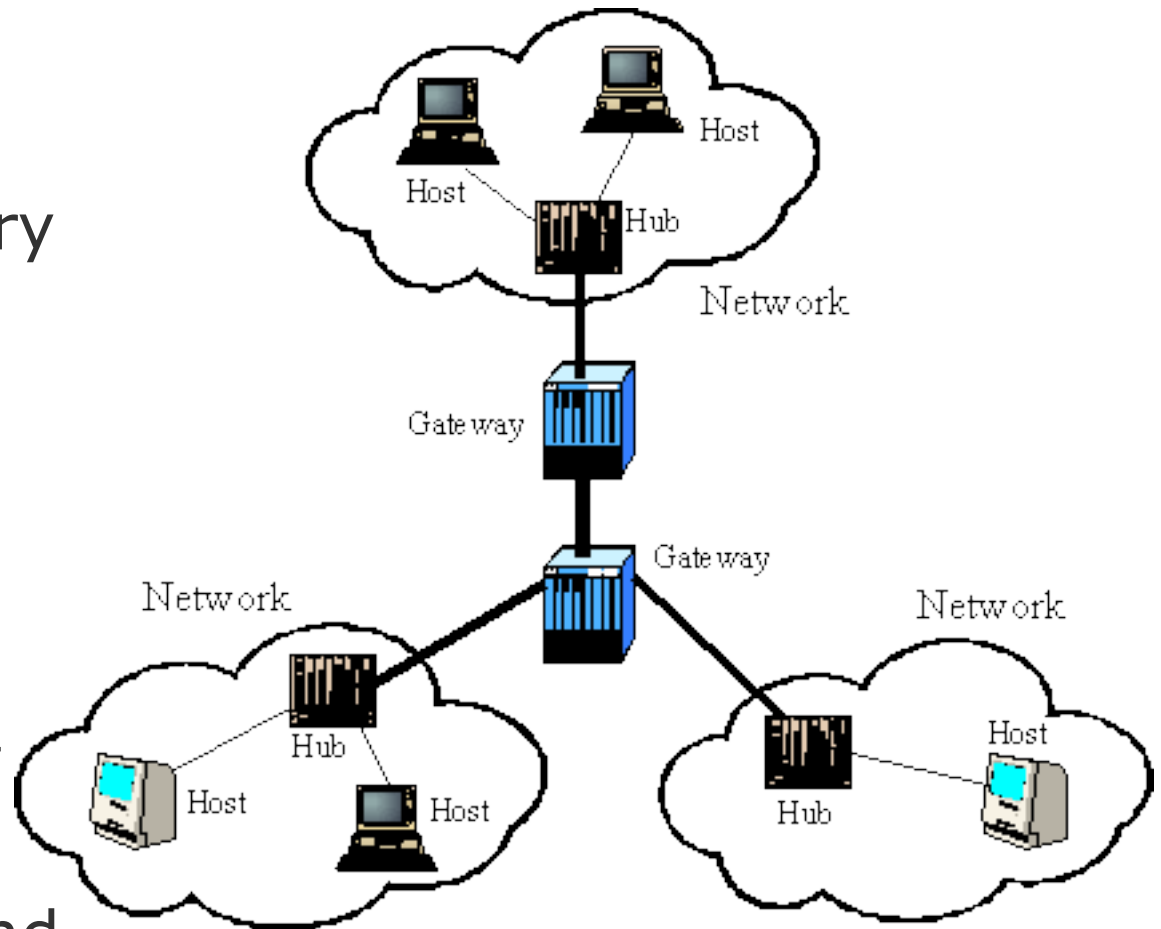
Network

Information is a network of binary relations.

Hierarchy is one kind of relation.

Taxonomies, Classification Systems are specific kinds of networks.

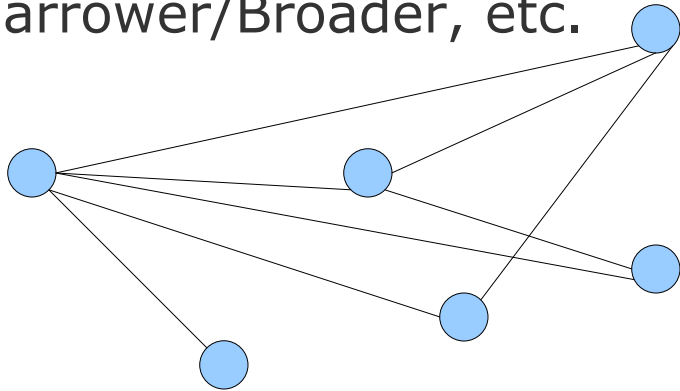
Internet is one kind of network.



<http://www.uga.edu/~ucns/lans/tcpipsem/internet.diagram.gif>

Network = Graph

- Graph = Nodes + Arcs
- Node
 - Atom, Account, Term, Subject, Person, etc.
- Arc
 - Composition, Naming, Typing, Genealogy, Narrower/Broader, etc.



Where does the Data Projection Model comes from?

- Topic Maps
- Resource Description Framework

Topic Maps

An ISO standard (ISO/IEC 13250)

Network of subjects

Generalized Connectivity

The Data Projection Model

has no specific semantics (topics, names, occurrences, associations, scopes, roles, etc.)

Resource Description Framework

Foundation of the Semantic Web (W3C)

Binary Relations:

- Generalized Triple Model (subject, object, predicate)

The Data Projection Model

- Has no specific semantics (description, title, etc.)
- Doesn't require to express information items as a URL.

Examples of Use

Maintenance of a Classification System

Maintenance of a Taxonomy

Maintenance of an Ontology

Maintenance of a Topic Map

Querying details within an information system.

Making explicit things that are implicit.

How to Use the Data Projection Model?

Integrating information from various sources

Enabling Multiple Concurrent Perspectives

1. Decompose into binary relations
2. Rebuild views according to biased perspectives.

Auditing Information Sources

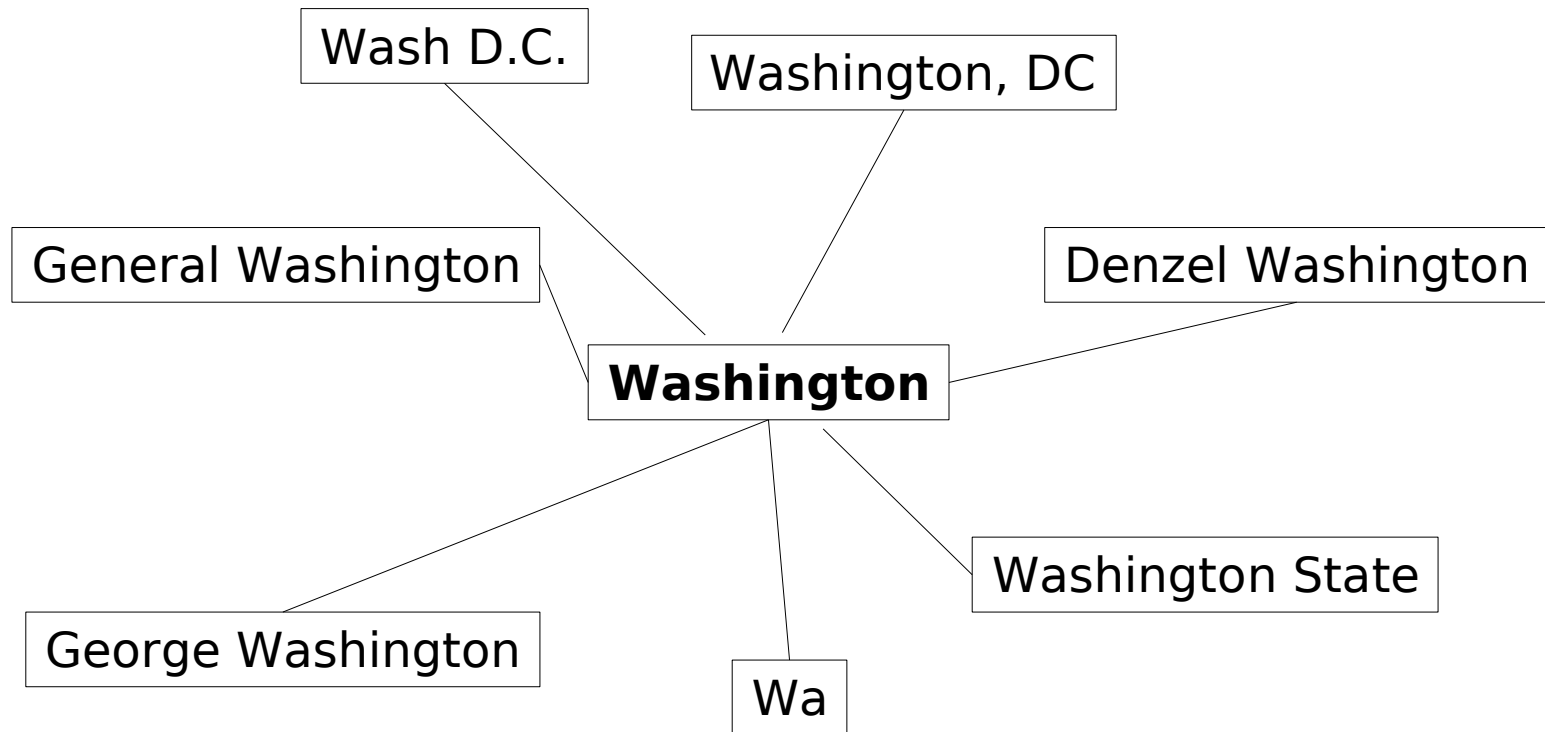
1. Auditing is a particular way of viewing things.
2. Can be used for explaining what happens, for quality control, etc.

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.

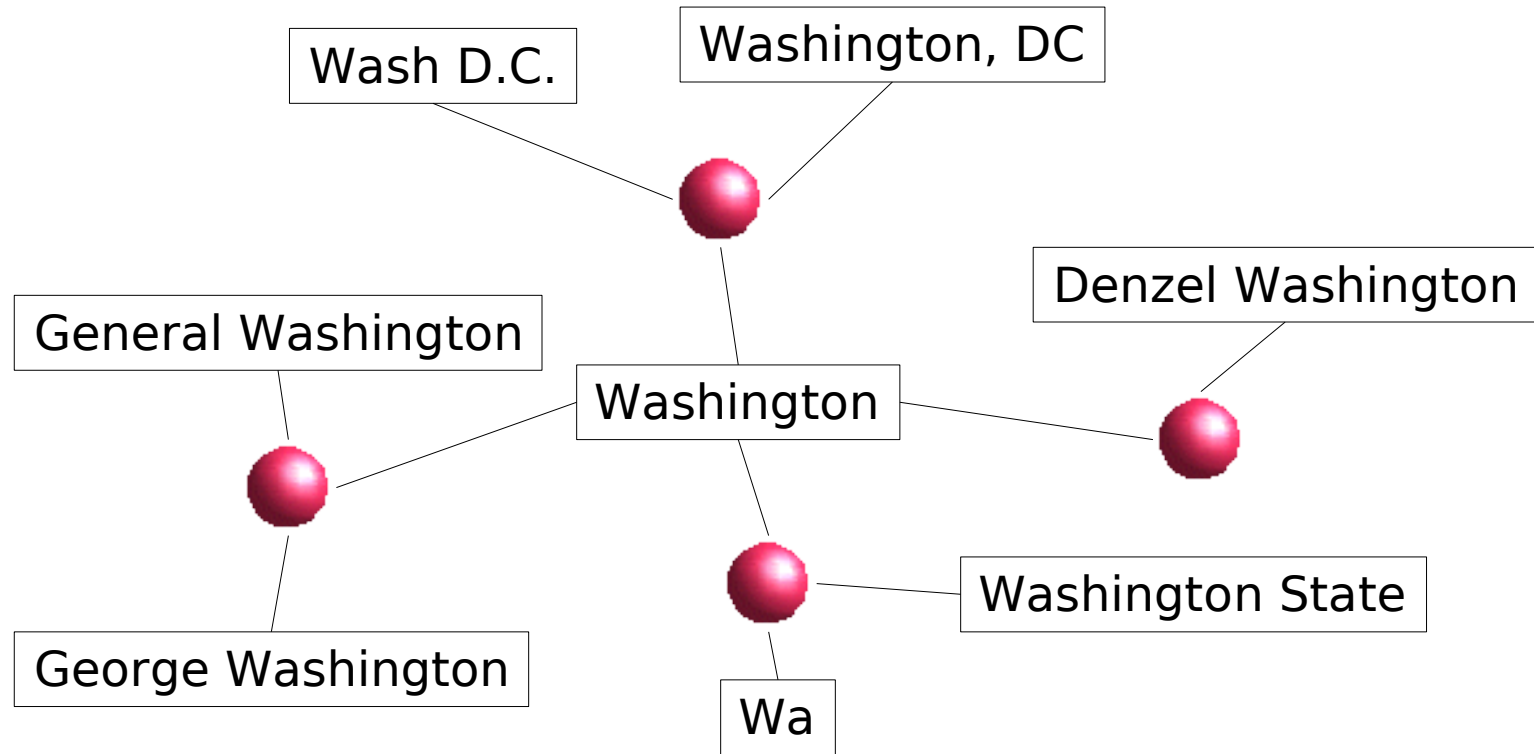
Names



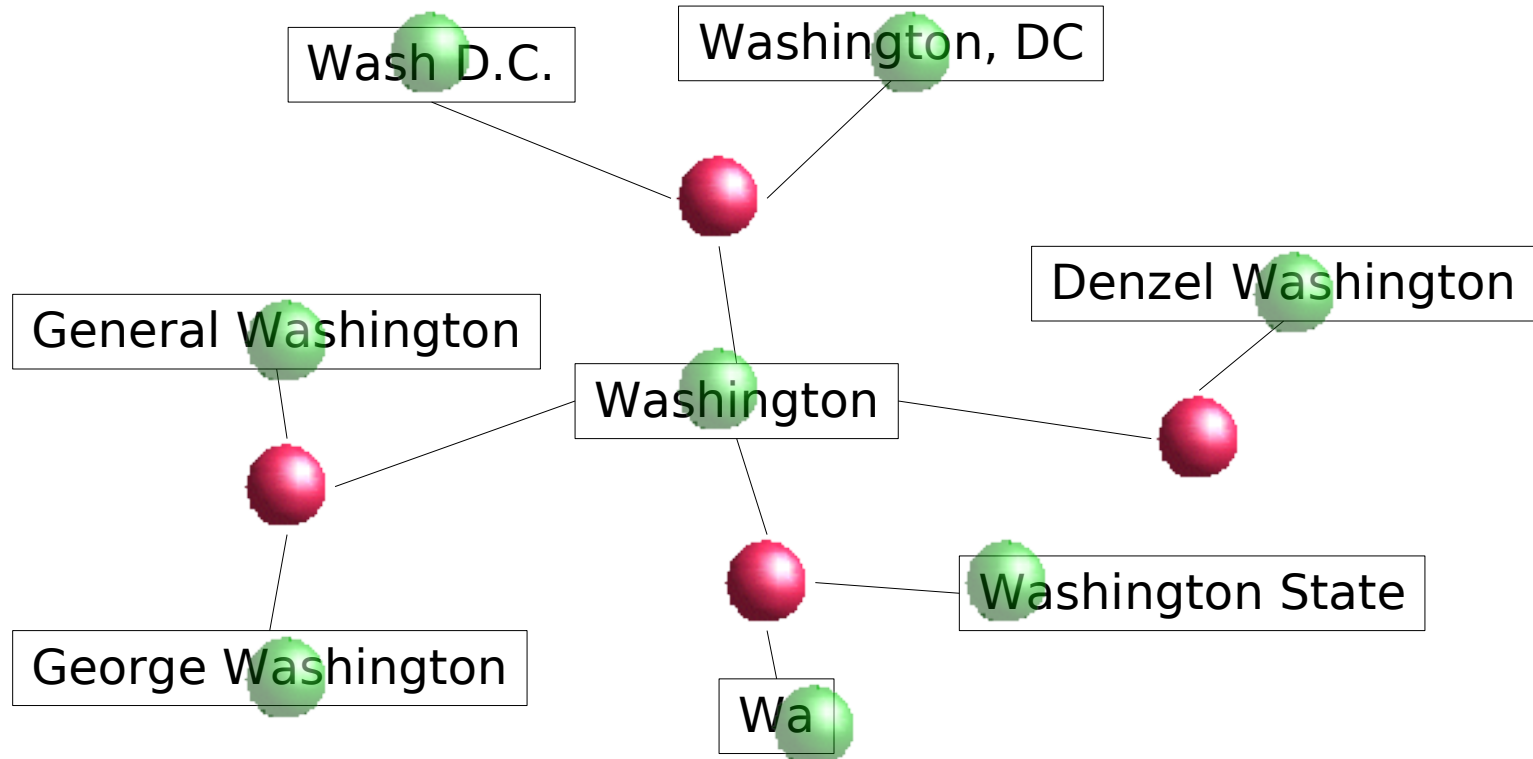
Names

< 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 >

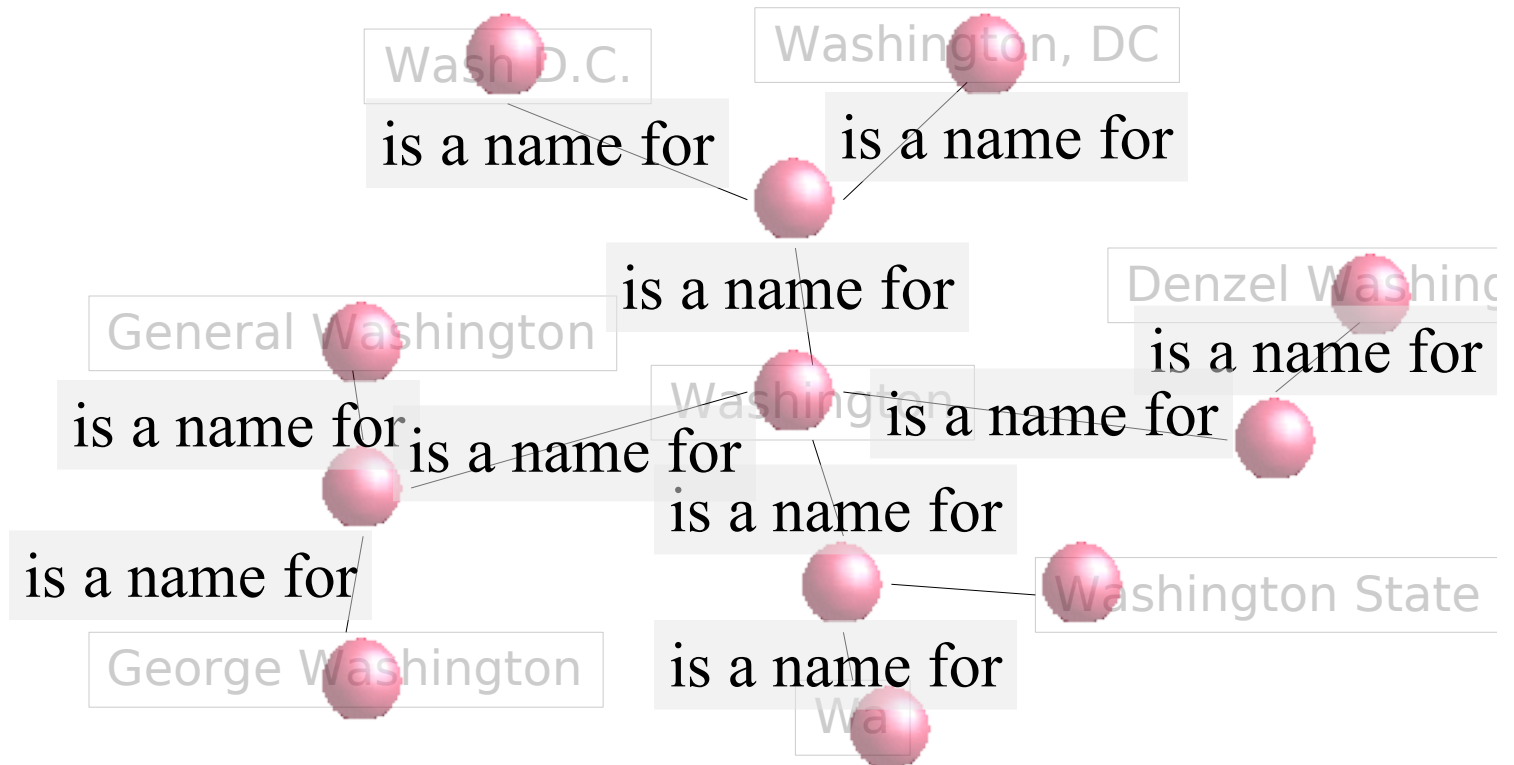
Emerging Subjects



Strings Become Subjects



Generalization



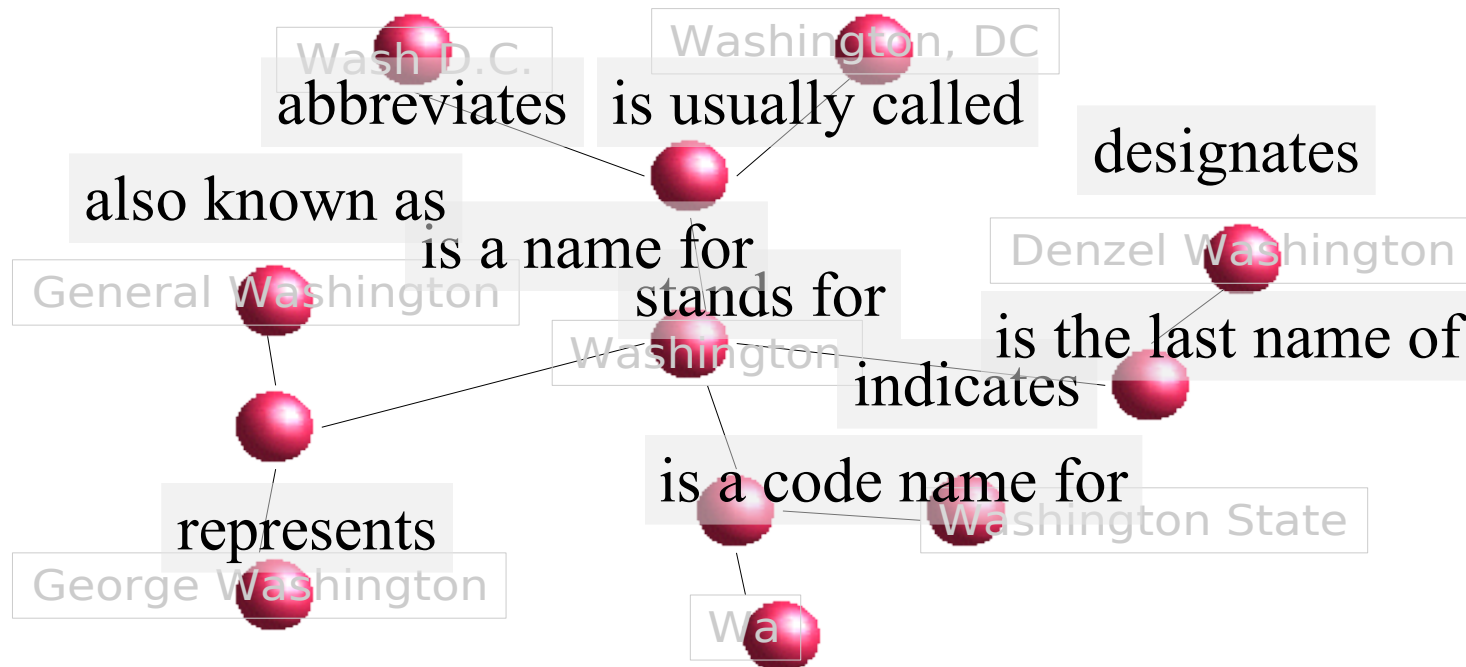
Names and Subjects

< 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	>

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	>
< 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	>
< 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.

Auditing

Auditing

- Accounting:
 - Single-Entry Bookkeeping:
 - Income: List of all we get that contributes to income.
 - Expenses: List of all our expenses.
 - Errors not detected. Records may be incomplete.
 - Double-Entry Bookkeeping:
 - Every transaction occurs between two accounts.
 - When one account gets credited, the other gets debited.
 - Checks and Balances. Accountability.

Information Accounting

- Double-Entry Information Accounting
 - No information item is ever isolated.
 - Transactions can describe processes (creation, deletion, etc.) or semantics (categorization, relatedness)
 - Each information item becomes an account that reveals all operations and connections ever made with it.
 - The Data Projection Model can be used for this.
- Details can be hidden from users.

Metadata, Data, and Projection

- The consideration of any piece of information either as data or metadata is a question of perspective...
- ... and many data can be both.

• Authors' Perspectives

- The Data Projection Model makes explicit the perspectives used by creators.
 - Highlight
 - Group

• Readers' Perspectives

- The Data Projection Model makes explicit the perspectives used to produce an output that is relevant to a given audience:
 - Filtering out
 - Presenting
 - Styles

Multiple Perspectives

- Multiple Perspectives can apply on the same set of data.
- Auditing view may be the most detailed view.
- End user views may be different from those of the original creators.

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



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](#)

Income Earned Abroad

[Publications](#)

[Publications](#)

[Links Inside Publications](#)

**Publication 593 - Tax Highlights for U.S.
Citizens and Residents
Income Earned Abroad**

[Income Earned Abroad You
exclusion from tax of a limi](#)

Related Topic Links

[Income](#)
[U.S. Citizens and Resident
Aliens Abroad](#)

Income Earned Abroad

Index

1	created	from document p593-001	>
2	related containment	Income	>
3	related PFT	Living Abroad	>
4	normalizes as	INCOME EARNED ABROAD	>
5	URL	ts0/tp6717.htm	>

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 decedents 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 name 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.infloom.com>

Michel Biezunski
Infloom

(718) 921-0901
mb@infloom.com