(The following is an excerpt from the Government of Alberta's paper entitled Privacy Architecture*, dated May 21, 2003, copyright ©2002, Government of Alberta. All rights reserved. Further information about the Government of Alberta's privacy architecture is available at* http://www.sharp.gov.ab.ca/ppa/*.)*

1. Privacy Taxonomy

This was originally referred to as the Privacy Classification Scheme but has been renamed as a taxonomy scheme to better reflect its intended use. Classification schemes, such as the Security Classification Scheme, are typically one-dimensional, simple, imply a fixed set of consequences and are mandatory to apply. A taxonomy scheme is more of an aid to breaking down a complex topic into categories that are useful to describe and specify anything relevant to the topic. Note that we still use the verb "to classify" to describe placing objects into the taxonomy!!

1.1.1 Design Points

The privacy taxonomy will:

- Be a means to allow other elements of the Privacy Architecture to be expressed in a compact and flexible manner, and to lead to the ability to make privacy-relevant decisions – automated or otherwise. (For example, "if classification is xxx then …")
- Not be an exercise for the sake of administration – each aspect of the scheme should directly support some other aspect of the Privacy Architecture or other architectures, such as the Security Architecture.
- Be used in whole or in part. It may not make sense to use the entire scheme everywhere, but to the extent it is used, the elements will always be consistent.
- Not be limited to data classification – any other parameters that are useful to classify to manage privacy in an IT environment will be considered.
- Be primarily for use with new initiatives. Decisions to retrofit for existing applications etc. will need to consider the costs vs. the benefits of applying the taxonomy.
- Position GoA for adoption of future privacy monitoring/enforcing technology
- Reconcile with recognized privacy classification/taxonomy schemes, in particular, P3P (and anticipate changes to those schemes).
- Adopt P3P-style acronyms for ease of interpretation
- Integrate with existing GAEA classification schemes (e.g. security)
- Include a starter set of GoA customized values for each category and an extensible set of codes for compact representation of the scheme
- Have a cross-enterprise core with ministry/program extensions

1.1.1.1 Rationale for Reference to P3P

P3P has emerged as the de facto standard for describing privacy policies in a format that can be interpreted by technology. P3P was developed by the World-Wide Web Consortium (W3C) and the full specification can be found at http://www.w3.org/TR/P3P/. P3P was designed to enable organizations to describe their web privacy policies to users versus enforcing those policies. As such, it naturally falls short of providing the additional vocabulary needed to describe the complexity of privacy within an organization required to enforce policy. For this reason, the recommended taxonomy is based on P3P but proposes significant extensions to it.

1.1.2 General Uses

The taxonomy scheme is designed to be used to whatever depth is appropriate to the situation. It may make sense to incorporate the scheme into the logic of new applications and the structure of new databases. On the other hand it may not be reasonable to retrofit existing applications and databases. In these cases it might serve only as an inventory or documentation aid. Here are some examples:

- It can be used to the full extent as a documentation aid to provide precise privacy-relevant descriptions of data stores (what kinds of PI they contain, the retention policy, the purposes it can be used for etc.).
- The taxonomy scheme will provide the language necessary to articulate a privacy policy for each PI Data Store in a structured, pseudo-code format. This will position GoA for future adoption of privacy technology (ex: P3P).
- The scheme can assist in developing privacy-specific audits of transaction records.
- The scheme can be used as a sanity check on proposed transactions. If the transaction cannot be expressed using taxonomy scheme values then the validity of the transaction should be questioned (the reverse is not true)

1.1.3 Taxonomy Scheme Overview

The basic proposal for a GoA privacy taxonomy is for a scheme with the following hierarchy:

- **Root Level** – contains “universal” dimensions that reference outside standards wherever possible
- **GoA Level** – contains GoA specific dimensions that will be common across GoA
- **Ministry Level** – contains Ministry-unique dimensions common within a Ministry

If dimensions are ever required below the Ministry level they can be defined at a later point.

Figure 1 shows the **Root Level** dimensions:

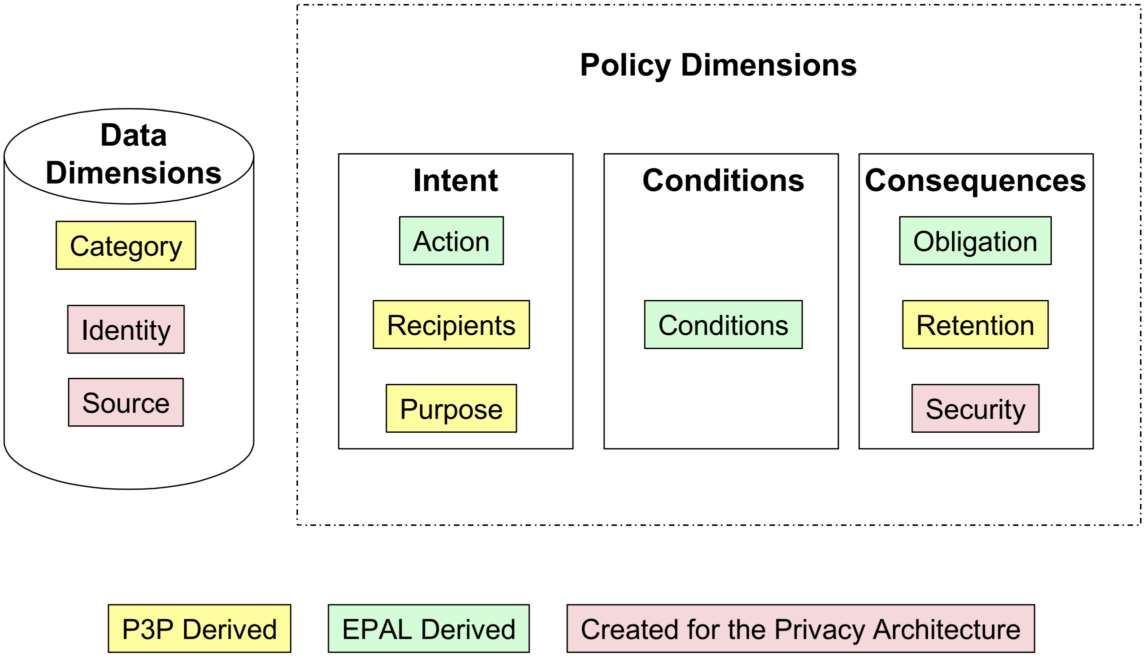

Figure 1: Privacy Taxonomy Root Level Dimensions

- Data Dimensions:
  - **Category,** ex: contact data, health data
  - ***Identity***, ex: personal information, anonymous information
  - ***Source***, ex: collected from the individual, derived, opinion
- Policy Dimensions (Intent):
  - ***Actions***, ex: collect, modify, use, transform, delete, disclose
  - **Purpose,** ex: provide health services, research, law enforcement
  - **Recipient,** ex: us or our agents
- Policy Dimensions (Conditions):
  - ***Conditions***, ex: require data subject consent, requires proof of authority
- Policy Dimensions (Consequences):
  - ***Obligations***, ex: inform data subject of right to appeal decision
  - **Retention**, ex: retain for purpose only
  - ***Security***, ex: the security level required to protect the information

The data dimensions represent characteristics of the data and do not change when policy changes (although a policy change may result in data classification needing to be more granular). The policy dimensions represent the parameters necessary to define a privacy policy (which will be defined against data as described by the data dimensions).

The dimensions in **regular type** represent existing P3P dimensions; the ones in ***italic type*** are proposed additional dimensions (the Actions, Conditions and Obligations come from the Enterprise Privacy Authorization Language (EPAL) specification).

The sections that follow will both describe in more detail the root level dimensions and will propose examples of GoA level extensions.

1.1.4 Data Dimensions

The data dimensions represent characteristics of the data and do not change when policy changes (although a policy change may result in data classification needing to be more granular).

1.1.4.1 Category

The basic proposal is to adopt the P3P broad PI categories as described in Table 1:

Table 1: Taxonomy, Category Dimension, Root Level

Code	Category	Meaning
PHY	Physical contact information	Information that allows an individual to be contacted or located in the physical world, such as a telephone number or postal address.
ONL	Online contact information	Information that allows an individual to be contacted or located on the Internet, such as an e-mail address. Often, this information is independent of the specific computer used to access the network.
UNI	Unique identifiers	Unique identifiers issued by a Web-site or service for the purpose of identifying an individual over time.
FIN	Financial information	Information about an individual's finances, including account status, account balance, payment or overdraft history, and information about an individual's purchase or use of financial instruments, including credit cards and debit cards.
DEM	Demographic data	Demographic and socioeconomic data, such as gender, age and income.
CNT	Content	The words and expressions contained in the body of a communication. For example, the text of an e-mail message, bulletin board postings or chat room communications.
PUR	Purchase information	Information generated by the purchase of a product or service, including information about the method of payment.
PRE	Preference data	Data about an individual's likes and dislikes, such as favorite color or musical tastes.
GOV	Government-issued identifiers	Identifiers issued by a government for purposes of identifying an individual over time, such as a driver's license number, social security number or passport number.
POL	Affiliation information	Information about membership in or affiliation with groups such as religious organizations, trade unions, professional associations, political parties, etc.
HEA	Health-related information	Information about an individual's physical or mental health, sexual orientation, use of or inquiry into health care services or products, and purchase of health care services or products.

Code	Category	Meaning
COM	Computer information	Information about the computer system that the individual is using to access the Internet, such as the IP number, domain name, browser type or operating system.
NAV	Navigation and click-stream data	Information generated by browsing the Web site, such as which pages are visited, and how long an individual stays on each page.
INT	Interactive data	Information generated from or reflecting explicit interactions with the Web site, such as queries to a search engine, or logs of account activity.
STA	State management mechanisms	Mechanisms, such as HTTP cookies, for maintaining an active connection with an individual or for automatically identifying an individual who has visited a particular site or previously accessed particular content.
LOC	Current Location Data	Information that can be used to identify an individual's current location and track them as their location changes – such as GPS position data

An example of extending the P3P scheme could be to add GoA level definitions to further sub-divide Health-related information into the three types defined in the HIA and assign new codes:

- Diagnostic Treatment and Care Information (DTC)
- Registration Information (REG)
- Health Services Provider Information (HSP)

Another extension could be sub-categories for Government Identifiers:

- Social Insurance Number (SIN)
- Alberta Driver's License Number (DLN)
- Health Unique Lifetime Identifier (ULI) – 9 digit numeric
- Child Welfare Identification Number (CWN) – 10 digit numeric
- Government Employee Identifier (GEI) – 10 digit numeric
- Alberta Student Number (ASN) – 9 digit numeric

The last four examples are recommended identifiers found under data standards on SHARP

As Electronic Service Delivery proliferates, offering choices regarding the way service is delivered is likely to become prevalent and so an individual's preferences will need to be captured. It is recommended that GoA be pro-active in this regard and define a common set of preference values. The PRE Category can be used to do this, for example:

- Language of choice (PLG)
- Preferred contact method (PCM)
- Preferred contact time (PCT)
- Preferred payment method (PPM)
- Web site customization preferences (favourite links etc.) (PCS)

Preferences, like contact information, is information that may be used across the GoA and may be a candidate for sharing.

Note that P3P does have existing levels of deeper granularity that can be used where appropriate. A good example is PHY (Physical Contact Information), which can be broken down into:

- Personal Address (PDR)
  - Street Address (STR)
  - City (CTY)
  - Province (PRV)
  - Postal Code (PST)
- Business Address (BDR)

(P3P does not actually assign 3-letter codes at this level, the ones above are just suggested).

The reason for showing the above as an example is that contact information is something of a special case both from the perspective of whether it is considered an identifier or not, and from the perspective that it is most often the example cited in terms of sharing personal information across governments. This will be discussed under the identity taxonomy. Also this illustrates that it is not a problem to have a hierarchy within a level if it is useful.

1.1.4.2 Identity

A critical concept associated with personal information that is recognized in legislation is that of the degree to which information is identifiable to a unique individual. This suggests the taxonomy should include a means of describing the identity level of data. P3P does not provide this so the following Identity dimension is proposed at the Root level as shown in Table 2

Table 2: Taxonomy, Identity Dimension, Root Level

Code	Identity	Meaning
PI	Personal Information	Information about an individual that includes information that readily identifies the individual.
DED	De-Identified Information	Information about an individual where the identifiers have been removed but keys have been retained to allow identity to be re-attached under the appropriate circumstances
WEK	Weakly Anonymized Information	Information about an individual where any identifiers have been permanently removed and the remaining information **have not** been transformed to further mask the identity of the individuals
STR	Strongly Anonymized Information	Information about an individual where any identifiers have been permanently removed and the remaining information **have** been transformed to further mask the identity of the individuals
AGG	Aggregated Information	Non-identifying information about groups of individuals

The detailed definition and usage of these identity levels will be discussed under the Data Transformation Topic but there is one level of granularity under the PI category that is useful to define here:

- **Individual Identifier** (INI) provides a practical means for identifying an individual. We define three classes of these:
  - **Unique Identifier** (UNI) is an individual identifier that is unique to an individual, known by the individual and usually assigned to the individual, such as an employee serial number or a Social Insurance Number.
  - **Non-Unique Identifier** (NNI) is an individual identifier that is not necessarily unique to an individual and may need to be cross-referenced with other information about the individual to confirm unique identity. Examples would include name, street address etc.
  - **Hidden Identifier** (HDI) is an individual identifier that is unique to an individual but that is not known by the individual. They are usually Meaningless But Unique Numbers (MBUNs) which are generated for use by applications and databases versus human beings
- **Attribute Information** (ATI) is information that describes an individual, or is about an individual, but which cannot readily be used to uniquely identify that individual. This is personal information only when it is connected to an individual identifier. (Examples of context information are age, salary, and contact preferences). Note that if enough elements of contextual PI are associated together it is theoretically possible to triangulate on identity if combined with a public information source that includes individually identifying PI. Attribute Information is the same thing as De-Identified Data except that the former term is used to refer to the information when it is still part of a PI record and the latter when it has been severed.

In other words, PI is made up of one or more Identifiers (Unique or Non-Unique) and Attribute Information.

Note, HIA defines **identifying** and **non-identifying** information which are broader than the identity levels described above because they are defined in terms of identity being "readily ascertainable". This is subjective and is not synonymous with whether the information contains individual identifiers or not. Identifying and non-identifying information are better thought of as assessments of data at the different identity levels. For instance one could look at a set of de-identified data and depending on the remaining attributes could declare it either identifying or non-identifying.

1.1.4.3 Source

The main use for a Source dimension is for severing records in situations like requests for access. Severing is likely to be performed based on the source of information. For example, information originally provided by an individual should always be accessible to that individual in the future whereas a recorded medical opinion about that individual may not be. Table 3 shows the proposed root level values for the Source dimension.

Table 3: Taxonomy, Source Dimension, Root Level

Code	Source
INV	Information directly collected from the individual
PRX	Information collected from a proxy (parent, guardian etc.)
3PY	Information collected from a 3rd Party (ex: another public body)
OBS	Information collected by observation or diagnosis of the individual
OPN	An opinion or evaluation made by others
DRV	Information derived, calculated, inferred or extrapolated

1.1.4.4 Sensitivity

Consideration was initially given to try and assign a sensitivity level to data but analysis showed that sensitivity is dependent on so may contextual factors that it cannot be assigned to data by a simple rule. The decision was to only use the term “sensitivity” as a qualitative descriptor and to include Security as a Consequence dimension since it is a consequence once sensitivity is determined in any particular case:

Note also that P3P does not attempt to classify some categories of data more sensitive than others. However, some implementations, such as Microsoft’s Internet Explorer Version 6 (IE6), do overlay a sensitivity interpretation. In the case of IE6, action is taken to restrict or block the use of cookies under certain scenarios, but only if PHY, ONL, FIN, GOV categories are included (which implicitly defines these as sensitive categories).

1.1.5 Policy Dimensions (Intent)

The policy dimensions represent the parameters necessary to define a privacy policy (which will be defined against data as described by the data dimensions). The first sub-class of the policy dimensions describe the intent potential data users may have relative to the data and include:

- Action
- Recipients
- Purpose

1.1.5.1 Action

Actions describe what the potential Data User intends to do with the data. Actions describe all of the manipulations that can be performed on data that have privacy relevance. The list of values for the root level of the Action dimension shown in Table 4 is taken from the EPA Architecture.

Table 4: Taxonomy, Action Dimension, Root Level

Code	Meaning
COL	Collect personal information
MDF	Modify personal information
USE	Use personal information (for the stated purpose)
DID	De-Identify information (a PI transformation action)
RID	Re-Identify information (a PI transformation action)
ANN	Anonymize personal information (a PI transformation action)
DSP	Dispose of personal information
DIS	Disclose personal information

NOT	Notify the individual of a privacy relevant issue
GCN	Give consent
WCN	Withdraw consent
PRI	Provide the individual with access to their personal information (i.e. Private Access)

The main use for defining Actions is to support the authorization process and position GoA for future adoption of technology that monitors or enforces privacy.

1.1.5.2 Recipients

Recipients describe both the intended users of the data and also any parties to which the data may be disclosed. We adopt the P3P defined Recipients as shown in Table 5.

Table 5: Taxonomy, Recipients Dimension, Root Level

Code	Recipients	Meaning
OUR	Ourselves and our agents.	This Web site, entities for whom it is acting as an agent, and/or entities acting as its agent. An agent in this instance is defined as a third party that processes data only for the completion of the stated purpose, such as a shipping firm or printing service.
DEL	Delivery Services	Legal entities performing delivery services that may use data for purposes other than completion of the stated purpose.
OTR	Other organizations following different practices	Legal entities that are constrained by and accountable to this Web site, but may use the data in a way not specified in this Web site's practices.
SAM	Other organizations following our practices	Legal entities that have equivalent practices to this Web site.
UNR	Unrelated third parties	Legal entities whose data usage practices are not known by this Web site.
PUB	General Public	Public forums such as bulletin boards, public directories, or commercial CD-ROM directories.

The primary area for GoA level definitions is additional granularity under "Other organizations following our practices" based on the fact that most disclosures are to other public sector bodies subject to the same regulations as GoA. Examples would be:

- Alberta Health & Wellness (AHW)
  - …. and all the other Ministries
- Law Enforcement Agency (LAW)
- Privacy Commissioner's Office (PCO)
- Health Services Provider (HSP)
- Learning Providers (includes schools, school jurisdictions, and post secondary institutions)
- The Individual (IND) – i.e. for Privacy Access
- A Proxy – such as a parent or guardian (PRX)

Section 40 of the FOIP Act defines disclosures permitted without consent and can provide a list of some of the organizations disclosures can be made to which could be added to the OTR type.

If needed, more granularity could also be defined under "Ourselves and Our Agents":

- Employees authorized to process the transaction (EMP)
- An Affiliate (AFF)
- An Information Manager (INF)

1.1.5.3 Purpose

Purpose describes why the data is to be used (e.g. the business justification) and we adopt the P3P defined Purposes as shown in Table 6.

Table 6: Taxonomy, Purpose Dimension, Root Level

Code	Purpose	Meaning
CUR	Completion and support of the current activity.	Information may be used by the Web site to complete the activity for which it was provided, whether the activity is a one-time event, such as returning the results from a Web search, forwarding an e-mail message or placing an order, or a recurring event, such as providing a subscription service or allowing access to an online address book or electronic wallet.
ADM	Web site and system administration.	Information may be used for the technical support of the Web site and its computer system. For example, to process computer account information, to secure and maintain the site, or to verify Web site activity by the site or its agents.
DEV	Research and development.	Information may be used to enhance, evaluate, or otherwise review the Web site, service, product or market.
TAI	One-time tailoring.	Information may be used to tailor or modify the content or design of the Web site during a single visit to the site. For example, an online store might suggest other items for a visitor to purchase based on items he has already placed in his shopping basket.
PSA	Anonymous user analysis.	Information that is based upon a unique identifier but that cannot be linked to an individual may be used for research, analysis, and reporting, For example, the number of users within a zip code.
PSD	Anonymous user profiling and decision-making.	Information that is based upon a unique identifier but that cannot be linked to an individual may be used to make a decision that directly affects that individual. For example, an individual within a certain zip code is presented with advertisements for companies located in that same zip code.
CON	Contacting visitors for marketing of services or products	Information may be used to contact an individual, through a communications channel other than voice telephone, for the promotion of a product or service. This includes notifying visitors about updates to the Web site.
TEL	Telemarketing	Information may be used to contact the individual via voice telephone for promotion of a product or service.
IVA	Individual User Analysis	Information that can be linked to an individual may be used for research, analysis, and reporting. For example, data about the types of and price ranges of products an individual has looked at.
IVD	Individualized decision-making	Information that can be linked to an individual may be used to make a decision that directly affects that individual. For example, a Web Site might show an individual houses that are within her ability to purchase, regardless of the price range she has researched before.
HIS	Historical preservation	Information may be archived or stored for the purpose of preserving social history as governed by an existing law or policy.
OTP	Other Purposes	Other Uses: <will include whatever text is specified>

The two primary areas for GoA level definitions are additional granularity under “Completion and support of current activity” and “Other Purposes”. An example of GoA level definitions could be to further sub-divide “Completion and support of current activity” into specific GoA purposes such as:

- Provide Health Services (PHS)
- Provide Learning Services (PLS)

Another example could be to further sub-divide “Other Purposes” into specific GoA purposes such as:

- Law Enforcement (LAW)

The best source to populate a starter set of these GoA specific uses would be the list of core activities described in the GAEA Business Architecture.

1.1.6 Policy Dimensions (Conditions)

The second sub-class of the policy dimensions is “Conditions”:

1.1.6.1 Conditions

Conditions describe the criteria that must be met in order to allow the potential Data User to carry out their intent on the data in question. Some examples of conditions that appear in FOP/HIA legislation are included in Table 7

Table 7: Taxonomy, Conditions Dimension, Root Level

Code	Meaning
CST	Requires individual consent
GDN	Requires Parent or Guardian consent
AUT	Requires statement of Legal Authority

A list of commonly cited Legal Authorities would be a logical extension of AUT at the GoA level.

The main use for defining conditions is to describe authorization requirements for access to personal information and position GoA for future adoption of technology that monitors or enforces privacy.

1.1.7 Policy Dimensions (Consequences)

The third sub-class of the policy dimensions describe the consequences of allowing potential Data Users to carry out their intent on the data in question and they include:

- Obligations
- Retention
- Security

Note that Retention and Security are just special cases of obligations that are drawn out for clarity and P3P compatibility.

1.1.7.1 Obligations

Obligations are additional actions that the potential Data User must undertake as a result of being allowed to carry out their intent on the data in question. Obligations are different from conditions because they don't impact the decision to allow the potential Data User to carry out their intent, but they do specify concurrent or future obligations the Data User must carry out if permission is granted. Obligations are specific to an organization and we define the following Root Level obligations for GoA in Table 8.

Table 8: Taxonomy, Obligations Dimension, Root Level

Code	Meaning
CCO	Concurrent obligations
FTO	Future obligations

An example of a GoA level definition would be:

- Inform data subject of right to appeal decision

This list should be expanded as required over time. The main purpose for defining obligations is to describe authorization requirements for access to personal information and position GoA for future adoption of technology that monitors or enforces privacy.

1.1.7.2 Retention

Retention describes how long the data will be retained by the organization and is usually initially determined when the data is first collected. Retention is just a special case of an obligation. We adopt the P3P Retention definitions shown in Table 9.

Table 9: Taxonomy, Retention Dimension, Root Level

Code	Retention	Meaning
IND	Indefinitely	Information is retained for an indeterminate period of time.
NOR	For the current request or session only	Information is not retained longer than the single online interaction.
STP	For the stated collection purposes only	Information is retained to meet the stated purpose and discarded at the earliest time possible (Must provide explanation)
LEG	As required by applicable law	Information is retained beyond the time it takes to complete the stated purpose because of a legal requirement or liability. For example, a law may allow consumers to dispute transactions within a certain time frame, therefore a Web site may decide to keep a record of transactions. (Must provide explanation)
BUS	As determined by our business practices	Information is retained per the service provider's stated business practices. (Must provide explanation)

The primary area for GoA level definitions is additional granularity under LEG based on the fact that most retention requirements are specified in law.

1.1.7.3 Security

This dimension is simply the existing GAEA security classification and describes the required security zone for the information in question as shown in Table 10. It is classed as a Consequence dimension because analysis determined that at this point in time, there are no static rules that allow the security level to be determined by simply looking at data dimensions like category and identity. Determining the security levels does use these dimensions as input but requires a PIA (Privacy Impact Assessment) process to determine the security level in each case (see the chapter on Data Placement). Over time, if these PIA decisions are tracked, it may be possible to make security into a data dimension.

Table 10: Taxonomy, Security Dimension, Root Level

Code	Security	Meaning
RAC	Restricted	Access is specific to an individual and very limited
CAC	Confidential	Access is specific to a function, group or role
IAC	Internal Use	Access is available to those possessing an authenticated identity
PAC	Public	Access is unrestricted

1.1.8 Implementation Considerations

1.1.8.1 Usage Guidelines for the Data Dimensions

For the purposes of this discussion, suppose that personal information is stored in tables such as Table 11 with a format where the columns are identifiers and attributes and the rows represent the values of those identifiers and attributes for a given individual.

Table 11: Sample Personal Information Table

Name	SIN	Address	Age	Salary
Homer Simpson	123 456 789	12 My Street	47	$60K
Seymour Skinner	372 809 875	245 Elm Avenue	52	$92K
Ned Flanders	375 059 354	11 My Street	45	$85K

The data dimensions of the taxonomy can be applied at the table, column, row or cell level as required. If applied to the table or to columns it can be applied via an external “meta-data” table. In English this might say “this table contains information gathered directly from individuals” or “this table contains names, social insurance numbers, addresses, ages and salaries”. In the notation of the taxonomy it would be more compact, something like: “SRC=IND”, “CAT=NAM,SIN,ADD,AGE,SAL”.

If applied at the row or field (cell) level then columns need to be added to the table itself to accommodate the values of the taxonomy. For instance, if most data in a table was collected from the individuals directly but some was gathered from guardians (and if it was important to know this difference) then a column for “source” would have to be added to the table.

Table 12 illustrates the probable application of the taxonomy to data tables. Entries in bold type indicate the most probably scenarios.

Table 12: Taxonomy Application Level

	Category	Identity	Source	Implementation
Table	Would be appropriate if the table only contains one category of data.	**Identity level will generally be most useful when applied to the whole table**	**Data in a table will most likely come from the same type of source**	Meta-data
Column	**Probably the most common method of applying the Category dimension especially for tables with multiple attributes**	If columns are to be separately used it could be appropriate to apply identity level by column	Might occur in situations where a certain category of data cannot be collected from the individual themselves	Meta-data
Row	Unlikely	Unlikely	Might occur if “proxy” situations are common (guardians etc.)	Within the table
Field	Unlikely	Unlikely	Possible	Within the table

For homogeneous data, the data dimension taxonomy can be applied at the table (or database or even server) level.

The recommendation for efficiency is that initially the taxonomy only be applied at the table and column level via meta-data tables. Applying at row and cell levels may only make sense once active privacy demands it. The implication of this recommendation is that a table of heterogeneous data will need to be labelled by rolling up the levels and generalizing the content to the most conservative description:

- Category: a list of all the Categories in the Table
- Identity: the highest level of Identity of any data in the Table
- Source: a list of all the Sources in the Table

It is also recommended that the policy dimensions be included in the meta-data descriptions wherever possible. It should usually be easy to at least apply the Security and Retention values to a database. Purpose and Recipients may be harder but would be useful. The remaining Policy dimensions such as Conditions and Obligations may often not be possible without going below the meta-data level.

1.1.8.2 Notation Format and Uniqueness of Codes

In addition to defining a privacy taxonomy scheme, it would be useful to define and agree on a notation format so that classifications can be documented compactly and unambiguously. As already mentioned, adopting 3-character codes for all categories and sub-categories is consistent with P3P and provides a scheme that can be “read” with a bit of familiarity. (These codes are actually only used for P3P compact policies but the idea is still valid).

The second element of notation is how to associate these 3-character codes. One possibility is to define a structure that allows the codes for a single dimension to be strung together in a consistent way such as:

Root Level.GoA Level.Ministry Level

For example:

HEA.DTC.PRE

For Health Information that is Diagnostic Treatment and Care Information that is a Prescription for Drugs. This does not limit the number of levels that can be defined below the Root Level as a “dot” can be used to separate each new level.

A related consideration is how unique the 3 digits codes are. **It is recommended that the root level codes be kept unique at the root level** (which is not unreasonable because the number of root level values should be relatively small and static). However. it would be easy to run out of unique codes at the lower levels which will be more populous and dynamic, and keeping them unique would require a lot of administration.

If this recommendation is adopted, the only requirement to keep this notation unambiguous is that all levels above the level to be described are included whenever the taxonomy is applied. So in the example above, DTC.PRE or just PRE may not be sufficient to unambiguously describe the data.

This recommendation would allow the re-use of root level codes even at lower levels. Re-used root codes could be chosen to mean the same thing as their root level value or something different as required. A good example of this is Health Registration information (REG), which is a GoA level code under Health Information. Health Registration information actually contains contact information (PHY) which is normally a root level value. This could be written as:

HEA.REG.PHY

Even though the notation syntax above unambiguously defines values in the taxonomy, it is recommended that a dimension level identifier be an allowable option so that processing or understanding may be speeded up in some situations. For instance, if CAT is the identifier for the Category dimension, the example above could be written as:

CAT=HEA.REG.PHY

This saves a look-up through all of the root level tables for all dimensions to determine that HEA.REG.PHY is a Category. Suggested dimension identifiers are shown in Table 13.

Table 13: Taxonomy Dimension Identifiers

Code	Dimension
CAT	Category
IDL	Identity (level)
SRC	Source
ACT	Action
PRP	Purpose
REC	Recipient
CND	Condition
OBL	Obligation
RET	Retention

SEC	Security

Clearly, both dimension and root level identifiers have to be unique between themselves (but can be reused at lower levels).

1.1.8.3 Administration

Once a substantial starter set of taxonomy value has been established there should be little requirement for a cross-government administration since root level values should be very static and even GoA level value should not see much change. **It is recommended that the GAEA Vitality Process be used to maintain the top levels of the taxonomy and that the Privacy Framework Review Committee be the key review/approval body.** Individual ministries can put their own administration in place if they wish to extend the taxonomy for their own use.

Note that some of the P3P dimension values may not find common usage in GoA, but it would not be a good idea to drop them because this would break compatibility with P3P. Once the initial population of the taxonomy is complete, P3P values not commonly used can be flagged. This may be useful in a negative sense - if someone believes they need to use a value that is rarely/never used, they might question the legitimacy of collecting it. For instance, if a particular initiative wishes to use Current Location Data, referencing the taxonomy they would find that this category is not in common use and they might want to get a policy decision made before they use it.

1.1.8.4 P3P Adoption

If GoA adopts this taxonomy, P3P implementation would be very simple. Since all P3P-based dimensions are at the Root Level, all that is required is to roll classifications up to this Root Level. In other words, even if GoA has very detailed breakdowns for describing data categories, purpose, recipients and retention at the GoA and Ministry levels, these all resolve back to the broad P3P category at the Root Level. The only other dimension to add would be access.

1.1.8.5 Tools

- In order to make the use of this notation to apply the Taxonomy simpler, consider creating a simple tool (e.g. a Spreadsheet) that departments could use to quickly create specific meta-data files for describing database content and policy
- In the near term, metadata could be captured in “PI Datastores” extended using the CITE toolset
- In the longer term, consider a complete metadata management solution

1.1.8.6 Default Values

Although the different data dimensions must always be independent by definition, as we move to lower levels it is possible and useful to make a default association between some of the dimensions. In particular, it is useful to attach identity values to commonly used data sub-categories as this makes decomposition of PI for access/placement much simpler. Table 14 shows the suggested default values.

Table 14: Default Category-Identity Associations

Code	Dimension	Identity
NAM	Full Name (first, last, initial)	Identifier
STR	Street Address	Identifier
CTY	City	Attribute
PRV	Province	Attribute
PST	Postal Code	Attribute
HPH	Phone Number (Home)	Identifier
EML	Email Address	Identifier
SIN	Social Insurance Number	Identifier
GEN	Gender	Attribute
MAR	Marital Status	Attribute
DOB	Date of Birth (year, month and day)	Identifier
CIT	Citizenship Status	Attribute
ETH	Ethnic Origin	Attribute

Clearly, identifiers such as Social Insurance Numbers are Unique Identifiers. The determination as to whether other identifiers are Unique or Non-Unique may depend on context. It is suggested that any identifier, or combination of identifiers, that are actually used to uniquely identify an individual in practice be labelled as Unique Identifiers. (For example, some Programs use a combination of Name, Street Address and Date of Birth to uniquely identify an individual).

It is recommended that the subset of pre-defined defaults for personal data elements above be adopted as part of the GAEA data standard. The most important and immediate effect would be that identifiers would be flagged as such. That way it would be crystal clear to developers that the privacy architecture applies for data stores containing personal data elements. Also, the application of the data standard would require a consideration of whether or not a given data element stored personal information.