

Personal Name Data Standard Rationale

Draft Proposal

For Discussion Purposes Only

Version 0.1

"NEWORG"

2011 12 30

Contents

01 Introduction	4
01.01 Purpose	4
01.02 Scope	4
01.03 Authority	4
01.04 Status	4
01.05 Related Documents	5
01.06 Legend	5
01.07 Related Documents	5
01.08 <Editorial Notes	5
01.09 Copyright	5
02 Person Name Conventions	6
02.01 Background	6
02.02 Summary of Name Conventions	6
02.03 Complexity	8
03 Objectives	10
03.01 Genealogist Uses for a Name	10
03.02 User Requirements	10
04 Rationale for Major Design Issues	11
04.01 Representation and Identification of NameParts	11
04.01.01 Problem	11
04.01.02 Alternatives	11
04.01.03 Analysis	12
04.01.04 Conclusions	12
04.01.05 Recommendations	13
04.02 Representation of Multiple GivenNames	13
04.02.01 Problem	13
04.02.02 Requirements	13
04.02.03 Alternatives	13
04.02.04 Analysis	14
04.02.05 Conclusions	14
04.02.06 Recommendations	15
04.03 Patronymic, Matronymic or Mononymic	15
04.03.01 Problem	15
04.03.02 Requirements	15
04.03.03 Conclusions	15
04.03.04 Recommendations	15
04.04 FamiliarNames	15
04.04.01 Problem	15
04.04.02 Requirements	15
04.04.03 Conclusions	16
04.04.04 Recommendations	16
04.05 FormalUseName and InformalUseNames	16
04.05.01 Problem	16
04.06 Data Entry/Input	16
04.06.01 Problem	16
04.06.02 Requirements	16
04.06.03 Alternatives	16
04.06.04 Conclusions	16

04.06.05 Recommendations	17
05 Outstanding Issues	18
05.01 Terminology.....	18
05.02 Genealogy	18

01 Introduction

01.01 Purpose

This Data Standard Rationale covers Personal Name, hereinafter referred to as PersonName (a technical Synonym). Although this data standard attempts to encompass all cultures and languages throughout the world, it is not intended to be exhaustive, i.e., address every level and range of complexity that exists. Rather, it provides solutions to the most common cultural usages in a flexible generic extensible manner.

It is targeted specifically for genealogical use, but may be applicable for many other purposes. In addition, it is intended to provide useful input into the "NEWORG" specifications.

This data standard is needed as no other comprehensive data standard exists that addresses all cultures and languages of the world and has been approved by any genealogy organization and one is urgently needed to improve the quality and reliability of PersonName data.

01.02 Scope

The scope of this data standard (see Appendix 01 Logical Data Model) includes:

- PersonName primary table
 - entity
 - attributes
 - derived values
- PersonName attribute's domain tables
 - entity
 - attributes
- PersonName attribute's link tables
 - entity
 - attributes
- PersonName external tables
 - entity
 - attributes

This data standard does not include any details on its external tables that are linked to PersonName, i.e. Person, Place, LocaleParameter, Note and Citation tables, as they are specified in other Data Standards from "NEWORG".

This data standard primarily addresses the logical data model of the PersonName entity. It is based on the objective of obtain maximum data structure stability by using normalized data entities represented in tables. Any relaxation of the principle of normalization, if necessary, is deferred until subsequent development of the physical data model. This design approach is appropriate regardless of the implementation technology used.

01.03 Authority

This standard is being developed by the "NEWORG" Project Team. Currently "NEWORG" is not working with any genealogical or technological organization on this data standard. ["NEWORG"](#) would welcome the opportunity to do so.

01.04 Status

The status of this standard is a draft proposal. Constructive critiques and suggested improvements are welcome and can be sent by E-Mail to [Neil Parker](#) or ["NEWORG"](#).

01.05 Related Documents

This document is based on "Personal Name Data Standard" by "[NEWORG](#)" that contains the background, name conventions and objectives. In addition, for each major design issue, it contains problem, requirements, alternatives, analysis, conclusions and recommendations for the Personal Name Data Standard.

The physical model is briefly addressed in a separate document, "[Personal Name Implementation Guidelines](#)" by "NEWORG". It contains implementation guidelines for software developers to facilitate increased data structure stability, data extensibility, data integrity, data accuracy and program maintainability.

01.06 Legend

Recommended PersonName terms are capitalized and concatenated, e.g., PersonName, Name and FamilyName.

01.07 Related Documents

This document is based on "Person Name Data Standard" By "NEWORG" that contains the background, objectives and requirements and, for each major design issue: problem, requirements, alternative, analysis, conclusions and recommendations for the Person Name Data Standard. In addition, this document contains a glossary that is applicable to this document.

01.08 <Editorial Notes

Editorial notes are shown in enclosed angle brackets and are meant to indicate:

- a comment by writer or reviewers
- an outstanding issue raise by the writer or a reviewer

Editorial notes will be deleted in subsequent versions.>

01.09 Copyright

<Boilerplate to be provided by "NEWORG">

This document is copywrited by "[NEWORG](#)". Users are free to use its contents provided they provide an acknowledgement that the data standard is the property of and is copywrited by "[NEWORG](#)". No commercial use may be made of this document from resale or fees.

02 Person Name Conventions

02.01 Background

The summary of personal name conventions below was abstracted from "Personal names around the world" with credits to original author (Richard Ishida) plus other information available on Wikipedia under Personal Names (see hyperlinks below):

<http://www.w3.org/International/questions/qa-personal-names>

[comments on this article](#)

[http://\"NEWORG\".wikispaces.com/file/view/gedcom55.pdf](http://\)

http://en.wikipedia.org/wiki/Personal_names

02.02 Summary of Name Conventions

The following is a summary of Naming Conventions used throughout the world.

FamilyName

- 01 The use of a FamilyName in Western culture is a relatively recent concept, beginning between the 12th and 19th century depending upon culture. Prior to these dates, the use of a patronymic (or a matronymic) as last GivenName was common and is still in use in some cultures today.
- 02 A mononymic is not a FamilyName.

SimpleName

- 03 Some cultures use a single letter for a name.

SingleInitialName

- 04 Many cultures use and encourage the use of an initial (capital letter followed by a period) instead of the middle GivenNames and, in some cases, the first GivenName.

PrefixedName

- 05 The SimpleName in a FamilyName and sometimes a GivenName may be prefixed by one or more short words; often conveying the meaning of the relationship between the person and a parent, place, or occupation. A space is used to separate these words with each other and the actual SimpleName.

CompoundName

- 06 There are cultural differences in the use of a CompoundName, especially for a FamilyName. Some cultures encourage/require the use of a CompoundName (often two NameParts but sometimes four NameParts) whereas other cultures discourage or deny their use. A CompoundName can be shown joined by a hyphen, nothing or a space.

GenerationalName

- 07 The Chinese often use a short word to qualify the GivenName of each new generation in a family. When used, it is always used as the first GivenName.

GenerationalTitle

- 08 Some Western cultures optional use Jr., Sr., I, II, III, First, Second, Third etc. to differentiate two family members in different generations having the name.

Designations

- 09 There are cultural difference in the use of designations, academic (e.g. BA, MBA, PhD) and professional (e.g. MD, PE, Atty) with some cultures treating their omission as a major error in etiquette.

NamePieceOrder

- 10 There are cultural difference in the order of naming components of Title, GivenNames and a Family Name for displays, printing and communication. [Hopefully it is the same for presentation and verbal/written communications.] Western cultures use GivenNames then FamilyName order whereas Eastern cultures like East Asia (e.g. mainland China, Japan, Korea, Malaysian Chinese, Singapore, Taiwan, and Vietnam), parts of Africa and Hungary use a FamilyName then GivenNames order.
- 11 There are cultural differences in where the Title is placed, i.e. at the beginning or the end of the above names.
- 12 There may be cultural differences in where the GenerationName is placed, i.e. at the beginning or the end of the above names.
- 13 There may be cultural differences in where the GenerationalTitle is placed, i.e. at the beginning or the end of the above names.
- 14 There are cultural differences in where the Designations are placed, i.e. at the beginning or the end of the above names and these have changed over time.
- 15 There are cultural differences in where to embed one or more FamiliarName among the GivenNames in a presentation of displays, reports or charts or FormalAddressingName and InformalAddressingName
- 16 There is no uniform convention for addressing a person verbally and in writing. Most cultures have a FormalAddressingName and an InformalAddressingName.
- 17 InformalAddressingName is often, but not consistently, used when the two people have been previously introduced and are known to each other.
- 18 The consequences of misuse of the FormalAddressingName or InformalAddressingName vary by culture, but some culture will consider their misuse poor etiquette if not an outright insult.

Indexing of GivenNames and FamilyName

- 19 There are cultural differences in the indexing of GivenNames and FamilyName as they appear in sequenced lists.
- 20 Most Western cultures index on the FamilyName, then GivenNames. Some cultures index on GivenNames, then FamilyNames. Indexing may vary depending on its purpose.
- 21 Many cultures use short prefixes in a PrefixedName for FamilyName. Sometimes these prefixes are part of the FamilyNameSortKey and other times not.
- 22 Some cultures use the first NamePart in a CompoundName in sorting Family Name, others use the last NamePart.
- 23 FamilyName is sometimes used as a last Name of GivenNames and therefore a GivenName (usually the last) can be any CompoundName. GivenNames also needs a separate GivenNamesSortKey e.g. "Jim" Walter James Smith.
- 24 Indexing of logogram names is often based on their pronunciation in a phonetic, Romantic or other language's alphabet. This process is perilous as logograms can have multiple pronunciations and accents. Several logograms can have the same pronunciation. Numerous errors are made in translating logogram to a phonetic, Romantic or other language.

Character Set Use

- 25 Western languages use letters whereas Eastern languages tend to use logograms.
- 26 Many Western cultures use a Roman alphabet

- 27 Other Western cultures using a Romantic Language use Roman letters with accents. Prior to Unicode an ANSI Latin 1 8 bit code, an accented letter was often implemented as a letter followed by the accent character.
- 28 Other Western cultures use a non-Romantic alphabet, e.g. Russian and Greece.
- 29 Each locale may have its own letter sort order, sometimes even differing among different countries using the same language.

Writing Direction of Names

- 30 Most Western cultures write left to right, then top to bottom.
- 31 Others Western countries (e.g. Hebrew and Arabic) write right to left, then top to bottom.
- 32 Eastern (e.g. Chinese, Japanese and Korean) logograms are usually written top to bottom, then left to right.

Name Component Terminology

- 33 There are difference is the terminology used to label the name components. In Western cultures, Given Names and FamilyName are the most common and useful term but they are sometimes replaced with first name or forename and last name which cause confusion to people using the FamilyName then GivenName NamePartOrder.
- 34 In Western cultures, the term GivenNames is sometimes replaced by Christian names, hardly cultural neutral, especial for many non-Christians.
- 35 In Western cultures, the term surname or last name is sometimes used, usually as a synonym for FamilyName.

There are many other cultural differences in names. Fortunately, many of these are concerned with the creation, derivation and formation of names rather than with their presentation and transfer so, although important to a genealogist, they are not pertinent to this data standard.

Cultural differences are not uniformly applied over time, over country, over language or even among the country's populous.

02.03 Complexity

A PersonName data standard is complex because of the following factors:

- temporal nature of the data
- usage depends on:
 - locale (language and country)
 - local customs, preferences and prejudices
 - individual preferences and prejudices
- lack of confidence in the data quality

Temporal implies that the use of a PersonName can change over time due to marriage, divorce, adoption, legal name change, use of aliases or "also known as" names, immigration (especially to a county speaking a different language), academic or professional certifications etc. Furthermore, these date ranges may overlap, i.e., at any given time, a person could be using two or more alternative PersonNames.

Usage depends on locale (language and country) and local customs (see Personal Name Data Standard Rationale, Section 02.02 Summary of Name Conventions). Regardless of culture, individual preference and prejudices often determines the initial giving of names and the subsequent use by the individual.

Historic source records are notorious for inaccuracies in recording what was said to the recorder if the provider of the information was illiterate or semi-illiterate in his native language or in a new countries a foreign language.

Most importantly, the past, current and future use of names is not within the control of the genealogist as they are to a much greater degree with typical business systems.

To be accepted in the marketplace, any good PersonName data standard must address these complexities. Adapting to this complexity requires compromises; it is not possible to design a perfect data standard in the face this complexity and lack of control over the input data.

03 Objectives

03.01 Genealogist Uses for a Name

A clear understanding of how genealogists use names is a prerequisite to determining the user requirements. Some of these uses are as follows:

- Identification of a person
- Checking with a person's ancestors' or descendants' PersonName for Name equality or similarity, especially when the Name may not be accurate, legible or complete
- Indexing list by FamilyName and GivenNames or each GivenName to assist in finding a person or other members of this person's family
- Presentation of display, reports and charts for examination and review by others
- For addressing the person with verbal/written communication, formally or informally

03.02 User Requirements

The following are required by this Person Name Data Standard:

- to determine the type of name.
- to respect all cultures and ensure the absence of culture bias.
- to accommodate the Name Conventions of various cultures throughout the world (see Section 02).
- to accurately determine NameParts, especially PrefixedNames and CompoundNames separated by a space.
- to sort GivenNames and FamilyName meaningfully according to user specifications.
- to recognize that a PersonName may only have been used for part of a person's lifetime and/or at a certain place.
- to address a person formally and informally for mail, e-mail or verbal conversation.

To achieve the above, this name data standard treats Names as an entity with primary and secondary attributes.

04 Rationale for Major Design Issues

Designing a data standard often involves compromises and tradeoffs, there is no perfect solution. Genealogy is particularly prone to this phenomena as historically there has been no standards or controls in place. People have been free to do what they please, both in the original recording of the data and in the subsequent inputting it into their genealogy databases. This situation still exist today and will likely continue in the foreseeable future. This problem becomes very apparent if one review the vast amount of data in old genealogy records file with LDS's Family History Records. These are based on evidence, based on separate individual recordings and not based on documented research based on evidence-conclusions. Hence much of the work is wrong, suspect, partially complete or has some other errors in it.

04.01 Representation and Identification of NameParts

This section is concerned with the representation and identification of:

- a PrefixedName
- a CompoundName joined by a space
- a GivenName separated from other GivenNames and/or Family Names by a space
- a Family Name at the front or back of a PersonName

04.01.01 Problem

A PrefixedName consist of a SimpleName prefaced by one or more words joined by spaces. Identifying a PrefixedName separated by its prefixes is a major problem because it is not possible to determine accurately if it is two or more NameParts or a PrefixedName. This problem is exacerbated if the PrefixedName is mixed with another NamePart or NameParts (i.e. SimpleNames, PrefixedNames and/or CompoundNames) as is possible in the GivenNames field or the FamilyName field.

A CompoundName may be formed by joining a NamePart or a PrefixedName with a hyphen, nothing or a space, e.g. Mary-Ann, MaryAnn or Mary Ann. Identifying a CompoundName joined by a space is a major problem because it is not possible to accurately determine if it is two NameParts or a single CompoundName in GivenNames or a FamilyName.

GivenNames are traditionally separated by a space. There is no way to tell if two names are two separate GivenNames or one CompoundName, e.g. Mary Ann Smith in storage of a database that stores GivenNames as a string or on presentation in displays, reports or charts.

Family names occur in GivenNames then FamilyName order in most Western cultures; but in FamilyName then GivenNames order in most Eastern cultures. Most database system store FamilyName in a separate field so this is not a problem; but when presented on displays, reports and charts, it may be impossible to tell which Name is the FamilyName.

These four aspects of the problem are exacerbated when presenting them on displays, reports and charts if artificial delimiters are not introduced. Introducing artificial delimiters to avoid ambiguity, may not look the same as people are accustomed to seeing it and, in fact, may look somewhat unaesthetic, depending on your tastes.

No known genealogy software or GEDCOM 5.5 addresses and provides a solution for the CompoundName problem nor the GivenName problem. GEDCOM 5.5 may provide a solution or partial solution to the PrefixedName problem as it provides for the use of a joiner between the prefixes, i.e. the comma (","). However, it appears this does not join the prefix(es) with the SimpleName word and therefore may not be sufficient in all cases. GEDCOM 5.5 provides a solution to the Family Name problem, i.e., enclosing it in slashes ("/").

Perhaps this problem has not been addressed in the past is because most historical records had no way of handling this issue unambiguously but rather rely on personal or local knowledge to visually make the appropriate interpretation. This is not as satisfactory and reliable solution for genealogical records.

04.01.02 Alternatives

01 Ignore the problem and accept that human interpretation is the best we can do.

- 02 Join each and every word including the SimpleName in a PrefixedName with one printable comma (","), join each NamePart in a CompoundName separated by a space in a GivenName or Family Name with one underscore ("_") and enclose FamilyName in slashes ("/").
- 03 Keep each and every word including the SimpleName in a PrefixedName joined by a space (" ") and each NamePart in a CompoundName separated with a space with a hyphen and terminate each Name in GivenName by a comma-space (", ").

Note: The joining character for a PrefixedName must be different that the joining character for a CompoundName in order to differentiate one from the other when they occur together, e.g. Smith van der Zam.

The following factors affect this decision:

- 01 Ability to convey to/from GEDCOM 5.5 files a PrefixedName and a CompoundName joined by a space for each GivenName and FamilyName
- 02 Ability to identify, process and display/print a PrefixedName and CompoundName joined by a space for GivenName and FamilyName
- 03 Ability to easily and accurately extract each Name in GivenNames or FamilyName.
- 04 Ability to market and convince professional and amateur genealogist that this Data Standard is a good approach.

04.01.03 Analysis

Obviously ignoring the problem is not addressing the problem. Some people may believe that the issue is not significant enough (occurs frequently enough) to warrant addressing. However other people believe that not being able to accurately identify and name a person by their NameParts is a major affront to the person and unprofessional genealogy.

These problem cannot be solved without using artificial delimiters (other than a space (" ")) with a different delimiter used between PrefixedNames, within CompoundNames, between GivenNames and around FamilyNames for presentation of displays, reports and charts. The NamePart problem can easily be solved for import/export from/to "NEWORG" but does not work for Import from GEDCOM 5.5.

The alternative of using the comma and underscore suffer from the difficulty of extracting GivenNames and FamilyNames correctly from existing GENCOM 5.5 data files in that GENCOM 5.5 contains no way of identifying a PrefixedName or CompoundName joined by a space for GivenNames or a FamilyName. On the other hand, any improvement in the new standard will improve the future but not make the past any worse. It probably suffers from another weakness; can we get other professional or amateur genealogist to accept it.

The only way to solve the CompoundName joined by a space problem is to place a delimiter such as a comma and space (comma-space) between each GivenName and FamilyNames. Although separating FamilyNames in a separate field solves this problem for Family Names in GEDCOM 5.5, it does not address the problem on display, hence the need for one comma between each part of a PrefixedName and one underscores between each part of a CompoundName separated by a space.

Although the next Section (04.02 Representation of Multiple GivenNames) can avoid some of the problems of separating multiple Names in GivenNames in storage or in data transfer, none of the alternative in the next section solve the problem of presenting data in displays, reports or charts.

04.01.04 Conclusions

In addressing this problem we must accept that we cannot change all the historic records that have not addressed this issue nor will we have any effect on how people continue to write their name in the future. We can, however, provide a facility in "NEWORG" that addresses this problem and publish this standard in the hope that other jurisdictions will eventually address the issue.

The best way to address the issue without introducing artificial delimiters is for Name issuing authorities (Registrars of Birth Certificates) to not allow a PrefixedName or a CompoundName separated by a space or if allowed to only use

nothing or a hyphen for joining their words or NameParts. This approach is naivety in the extreme and will be difficult, if not impossible to achieve, in a democratic society.

Both Alternative 2 and 3 appear to be able to meet the requirements with different implementation complexities but about the same magnitude. Unfortunately, both alternatives change the presentation of the PersonName for displays, reports and charts. One could argue that as PrefixedNames and CompoundNames occur relatively infrequently, that this problem is not that significant. [What do you think? What will others amateur and professional genealogist think?]

04.01.05 Recommendations

- 01 Use one comma (",") to join each and every word including SimpleName in a PrefixedName joined by a space in Given Names or Family Names similar to the way that GEDCOM 5.5 recommends.
- 02 Use one underscores ("_") to join NameParts in a CompoundName joined by a space in GivenNames or FamilyNames'
- 03 Separate GivenNames by a space and enclose Family Name in slashes (solidus).

04.02 Representation of Multiple GivenNames

Assuming that we accept the preceding recommendations, the following problem arises.

04.02.01 Problem

GivenNames can have one or more GivenNames with a total character length of one or more characters. Although some jurisdictions responsible for birth registration now place a maximum limit on the overall length, one cannot assume any specific maximum limit on historic data, either in the number of GivenNames or in their total character length.

[P.S. GEDCOM 5.5 is full of artificial limits on field lengths; some may be unnecessary. Each should be examined for necessity and reasonableness.]

04.02.02 Requirements

- 01 A PrefixedName must be identifiable and be capable of being treated as a single GivenName
- 02 A CompoundName must be identifiable and be capable of being treated as a single GivenName.
- 03 GivenName order is significant, cultural sensitive and needs to be preserved.
- 04 Genealogy and/or many other applications have a need to identify the first GivenName and middle GivenName.
- 05 Genealogy and/or many other applications have a need to identify the middle GivenNames for complete identification.
- 06 GivenNames and each GivenName need to be indexed to support finding and selecting persons and identifying similarly named potential family members.
- 07 A mononymic GivenName must be identifiable but treated as a first GivenName.
- 08 FamiliarNames are usually placed at the end of all GivenNames although some cultures or people like to place them after the specific GivenName that they represent, e.g., Elizabeth Margaret "Maggie" Georgina Smith, this seem to be primarily a matter of individual preference.

04.02.03 Alternatives

Given names could be stored and exchanged in:

- 01 one field labeled and containing all GivenNames, each Name separated by one and only one space (" ").
- 02 one field labeled and containing zero or one FirstGivenName and a second field labeled and containing zero or more MiddleGivenNames, each Name separated by one and only one space (" ").

- 03 multiple fields (of a specified maximum occurrence), labeled (i.e., GivenName1, GivenName2...) and containing zero or one GivenName; once a field contains zero Given Names, all subsequent fields must contain zero GivenNames
- 04 a separate entity in a GivenName domain table with a link table showing the many to many relationship between PersonName and GivenName (i.e. the fully normalize approach).

The following factors affect this decision:

- 01 Ability to implement referential integrity on each GivenName in GivenNames (if considered desirable).
- 02 Ability to index all Names in GivenNames together
- 03 Ability to index each GivenName separately (if indexes are required on each individual GivenName).
- 04 Ability to convey to/from GEDCOM 5.5 files a PrefixedName and CompoundName separated by a space.
- 05 Ability to identify, process and present displays, reports and charts of PrefixedNames and CompoundNames separated by a space.
- 06 Ability to determine the GivenNames, first GivenName, middle GivenName, MiddleGivenNames and each GivenName.

04.02.04 Analysis

All alternatives suffer from the difficulty of extracting given names from existing GEDCOM 5.5 data files in that GEDCOM 5.5 does not have a method of identifying a CompoundName separated by a space. On the other hand, any improvement in the new Person Name Data Standard will improve the future but not make the past any worse.

Regardless of which alternative is selected, there still needs to be a method of identifying PrefixedNames and CompoundNames on presentations of displays, reports or charts. This requirement is difficult to achieve in presentation on displays, reports or charts that appears as a continuous string of GivenNames without introducing some artificial joiner character for each PrefixedName or CompoundName such as a comma(",") for a PrefixedName and an underscore("_") for a CompoundNames. Although this approach is somewhat unaesthetic, it appears to be the best alternative (least unaesthetic) available. This approach should be implemented in all four alternatives above on both internal storage and presentation and transfer.

Having separate fields as in Alternative 3 and 4 has the following

Advantages:

- implementing referential integrity on each GivenName using built-in DBMS features is easier and more reliable than the alternative of hand coding by the programmer
- Implementing indexing on each GivenName is easier, if this is a requirement
- Makes the issue of data identification and extraction from the database simpler by eliminating the need for a parsing function to extract each GivenNames from one field
- Alternative 4 places no limit on number of GivenNames or on their individual or overall length (in storage).

Disadvantages

- implementing indexing on all GivenNames is more difficult, however the GivenNamesSortKey provides the solution to this problem
- Alternative 3 requires specifying a maximum number of GivenNames fields with five being suggested as an arbitrary but a reasonable maximum limit although this could easily be increased.

04.02.05 Conclusions

All alternative methods appear to be able to meet the requirements with different implementation complexities but about the same magnitude providing that a standard way such as the use of a comma or a underscores to solve the PrefixedName or CompoundName joined by a space problem.

Assuming the recommendation is acceptable, then migrating to the normalized technique of placing GivenNames in a separate table as it allows the most functionality of any of the alternatives.

04.02.06 Recommendations

- 01 Design the data base so that GivenNames are place in their own domain table and use a link table to join them with the PersonName.
- 02 Because FamiliarNames also can occur many to many times, implement the same technique with FamiliarNames.
- 03 Because Designations also can occur many to many times, implement the same technique with Designations.

04.03 Patronymic, Matronymic or Mononymic

04.03.01 Problem

What special treatment does a patronymic, matronymic or mononymic require?

04.03.02 Requirements

Patronymic and matronymic are just special forms of GivenNames being derived from paternal or maternal ancestor's FamilyName or first Name of the GivenNames.

A mononymic, by definitions, must be the only PersonName attribute and it must appear in the first GivenName field.

04.03.03 Conclusions

No special treatment is required of patronymic or matronymic but perhaps a short comment could be entered in the Description or Note attributes indicating its identification and derivation.

Mononymic should have some edit checks undertaken to ensure that no other Name field contains data other than the first Name of GivenNames and that Type = Mononymic.

04.03.04 Recommendations

- 01 GivenNames may also include a patronymic or matronymic as a FamilyName or as the last GivenName of GivenNames.
- 02 Mononymic may only appear as the first GivenName of GivenNames and all other name fields must be null. The use of PersonNameType = Mononymic should be used to identify Mononymic and enforce this rule.

04.04 FamiliarNames

04.04.01 Problem

FamiliarNames need to be identified and placed as the last GivenName in the GivenNames field. The order of placing FamiliarNames may be culture sensitive, but often left to the users preference.

GEDCOM 5.5 provides a separate field for FamiliarNames but some genealogy software does not support this in their PersonName template, e.g. Legacy.

04.04.02 Requirements

FamiliarNames need to be identified as such so they are not mistaken for a real GivenName.

04.04.03 Conclusions

Identification can easily be achieved on presentation of displays, reports and charts by enclosing each FamiliarNames in double quotation marks(""), a convention that is pervasive today.

04.04.04 Recommendations

- 01 Embed each FamiliarName in presentation (not in storage) in double quotation marks to identify the FamiliarName.
- 02 Each and every FamiliarNames should appear at the end of GivenNames.
- 03 One or more FamiliarNames can never appear between the parts of a PrefixedName or CompoundName.
- 04 User should have the ability to suppress the presentation of FamiliarName in displays, printed reports or charts.

04.05 FormalUserName and InformalUseNames

04.05.01 Problem

Currently the Data Standard proposes having two names; a FormalUserName and an InformalUserName. While reflecting on this issue, it occurs to me that there are really more than two and that perhaps this attribute should be NameFormality in a separate domain table with its own link table to support the many to many relationships and supporting the following UseNameFormalities Types:

- Formal with Designations
- Formal
- Informal with Designations
- Informal
- Casual

Each Type could have an associated list of sequenced NamePieces as parameters and these parameters could be in the Locale table.

04.06 Data Entry/Input

04.06.01 Problem

Existing PersonName data contains many errors that this data standard attempts to address and that GEDCOM 5.5 does not provide a capability for. If "NEWORG" data standard required rejecting this data, it would not gain wide acceptance from the user community.

04.06.02 Requirements

Ability to accept all PersonName data "as is".

04.06.03 Alternatives

- 01 Accept all PersonName data "as is" but issue warnings of errors or possible errors during import.
- 02 Reject erroneous data.

04.06.04 Conclusions

Reject erroneous data will probably result in "NEWORG" not receiving the wide acceptance it needs to succeed.

04.06.05 Recommendations

- 01 Accept all input but issue warnings of errors or possible errors during import unless it's a simple correctable error, e.g. two or more spaces between NameParts or name components.
- 02 All genealogy software should have a user executable feature to audit the genealogy database and issue warning messages for any PersonName errors or possible error found and the reason for the error or possible error.

05 Outstanding Issues

05.01 Terminology

01 Are the Terms being used the most appropriate?

05.02 Genealogy

01 How does GEDCOM 5.5 or "NEWORG" implement many to many relationships, i.e. Link Tables, in addition to Notes and Citations, we have this issue in GivenNames, FamiliarNames and Designations? It already does this in the Family Entity, are there other examples?

02 Does FamiliarNames always occur just after GivenNames, if so, we do not need it in the NameOrder field?

03 Do we have to allow FamiliarNames always occur just after GivenNames? If so, each should be entered with its enclosing format QuotationMarks as a separate GivenName.

04 Does GenerationalTitle always occur just after FamilyNames, if so, we do not need it in the NameOrder field?

05 Is a Locale for PersonName Configuration attributes desirable with one record for each Locale (Language-Country) that contains parameters for characteristics such as:

- PrefixedNamesAllowed
- CompoundNamesAllowed
- CompoundNameJoinerCharactersAllowed
- NameOrder
- CharactersOrLogogramsAllowed
- LetterSortOrder
- Non-LetterCharactersAllowed
- FormalAddressingNameFormat
- InformalAddressNameFormat
- WritingDirection

06 Does PersonName need a Locale attribute?

07 Should the installation provide a DefaultLocale referencing the PersonName Locale Configuration attributes with a Locale value for each PersonName, if not null, overriding the installation default Locale value?

08 Do we need to only sort on the GivenNames or is it important to be able to index each and every GivenName? Genealogist, Police and others may want an index on each GivenName to aid in identifying someone? If so, why?

09 Do genealogist or others ever need to access just first GivenName singularly? If so, why?

10 Do genealogist or others ever need to access just middle GivenNames together? If so, why?

11 Do we need family name delimited by enclosed "/"? Why? Is here a better, more aesthetic way to achieve the same result.

12 How extensive a description should we give on the different views of data, i.e. presentation

GivenName treatment

- output format, determined by user setting for each of display, reports and charts, correspondence
 - as a printable comma-space separated values
 - as an printable underscore separated values
 - as a space separated values (at user's peril)
- output format for export is as defined by "NEWORG"
- input parse for data entry, a single comma-space separated value
- input parse for import
- from "NEWORG" is as defined by "NEWORG"
- from GEDCOM 5.5 with no delimiters assumes space separates each name

- from GEDCOM 5.5 with comma-space delimiter is no problem as defined by "NEWORG"

13 Is the horizontal ellipsis ("...") a better (less unaesthetic) symbol to use as a PrefixedName joiner than GEDCOM 5.5 caret ("^")? An ellipsis is three horizontal dots on the writing line representing one character. In Microsoft Office, it appears by default if you type three consecutive periods.