Explorations in a Family History Quality/Completeness Metric

Family History Technology Workshop

Brigham Young University

February 2, 2016

Ben Baker

bakerb@familysearch.org

FamilySearch

Background

- Family history researchers have good feel for completeness on a person

- Desktop record managers and online trees provide mechanisms to record information, but are missing a way to quantitatively discern the quality or completeness of a given person

- Built upon investigation by Kent Olsen and others at FamilySearch

- No working implementation yet – Still exploring and seeking feedback

Goals

Develop a quantitative metric

- To measure the quality and/or completeness of persons in a family tree
- To drive user behavior towards desirable tasks to improve data to be
  - More complete and well documented
  - Free of data problems, issues and duplication

FamilySearch

Possible uses

- Measuring the quality or completeness of a person record

- Suggesting areas for users to work to improve the data of their own relatives

- Introducing impedance to modification of very complete person records

- Providing guidance to improve results when merging duplicate persons

- Tracking overall quality improvements over time to the tree as a whole

- Share items from high quality persons – Ex. E-mail a story, show as an example

FamilySearch

Metric Main Components

Component	Weight (non-LDS)	Weight (LDS)
Information	20%	16%
Sources	30%	26%
Quality	35%	31%
Memories	15%	12%
LDS Temple Ordinances	N/A	15%

FamilySearch

Information

Presence and completeness of information. Will lose quality points for certain types of data that are often indicative of problems

- Presence of a name
- Complete name – Ex. has both first and last name pieces
- Clean name – Ex. No slashes, titles as first name, etc.
- Name seems to match name template
- Presence of a gender
- Presence of a life sketch
- Presence of a birth-like event with at least a date and/or place
- Birth-like date is more than year only. (assumes date was standardized)
- Birth-like place has more than one jurisdiction (assumes place was standardized)

- Presence of a death-like event with at least a date and/or place
- Death-like date is more than year only (assumes date was standardized)
- Death-like place has more than one jurisdiction (assumes place was standardized)
- Has at least one spouse or died before marriageable age
- Presence of a marriage event with at least a date and/or place
- Marriage date is more than year only
- Marriage place has more than one jurisdiction (assumes place was standardized)
- Has at least one father
- Has at least one mother
- 1 point for each other information event or fact to a maximum of 3 points

FamilySearch

Sources

Each attached source worth 1 point – maximum 10

Open questions:

- How to handle different time periods and areas of the world that affect how many sources there can be
- Is a structured source more valuable than a user contributed source?

2 points for each tagging

- Name is tagged with a source
- Gender is tagged with a source
- Birth-like event is tagged with a source
- Death-like event is tagged with a source

Quality

- Start with 30 points subtracting the following to a minimum of 0 points
    - 5 points for each data problem
    - 2 points for each record hint
    - 2 points for each research suggestions
    - 2 points for each possible duplicate
    - 2 points for each additional set of biological parents
    - 2 points for couples with more than 20 biological children
    - 1 point for 3+ alternate names

FamilySearch

Memories

- 3 points for a portrait

- 1 point for each memory (photo, document, story, audio) to a maximum of 10 points

- Need some way to handle availability based on time period and areas of the world

LDS Temple Ordinances

Points for each ordinance based on its status

0 – Needs More Information
1 – Ready
2 – In Progress
3 – Completed or Not Needed

Ordinances
Baptism/Confirmation
Initiatory
Endowment
Sealing to Parents
Sealing to Spouse

Final Score

- Scores in each category are weighted and normalized to a 0-100 scale
- For user presentation, each category and the overall metric are displayed as star rankings

Low	High	Stars
0.0	9.1	
9.1	18.2	★
18.2	27.3	★
27.3	36.4	★★
36.4	45.5	★★
45.5	54.5	★★★
54.5	63.6	★★★
63.6	72.7	★★★★
72.7	81.8	★★★★
81.8	90.9	★★★★★
90.9	100.0	★★★★★

FamilySearch

Few real examples

William Alvin Tolman - KWCF-RKJ

Category	Points	Possible	Percentage	Non-LDS Weight	LDS Weight	Star Ranking
Information	20	21	95%	19%	15%	★★★★★
Sources	18	18	100%	30%	26%	★★★★★
Quality	26	30	87%	30%	27%	★★★★⯪
Memories	13	13	100%	15%	12%	★★★★★
LDS Temple Ordinances	15	15	100%		15%	★★★★★
Overall				94%	95%	★★★★★

Ann Smith - L8QV-SHD

Category	Points	Possible	Percentage	Non-LDS Weight	LDS Weight	Star Ranking
Information	14	21	67%	13%	11%	★★★⯪
Sources	8	18	44%	13%	12%	★★
Quality	26	30	87%	30%	27%	★★★★⯪
Memories	0	13	0%	0%	0%	
LDS Temple Ordinances	12	15	80%		12%	★★★★
Overall				57%	61%	★★★

FamilySearch

Susie Christine Nolty - KH24-BCH

Category	Points	Possible	Percentage	Non-LDS Weight	LDS Weight	Star Ranking
Information	17	21	81%	16%	13%	★★★★
Sources	0	18	0%	0%	0%	
Quality	14	30	47%	16%	14%	★★⯪
Memories	0	13	0%	0%	0%	
LDS Temple Ordinances	15	15	100%		15%	★★★★★
Overall				33%	42%	★⯪ or ★★

FamilySearch

Future Plans

- Validate with genealogists and less experienced users, adjusting as appropriate

- Work with FamilySearch research team to assess applicability across entire tree

- Resolve or work around performance issues, particularly with possible duplicates and LDS ordinances

- Examine impact of popularity factors (Ex. Number of views/likes, watchers, edits, unique contributors, reverts, report abuse incidents etc.)

FamilySearch