Constructing a family tree in an online space as an act of digital curation: An auto-hermeneutics study

Donovan Stuart Porter

PRTDON002

Supervisor: Michelle Kahn

A minor dissertation submitted in partial fulfilment

of the requirements of the degree of

Master of Philosophy in Digital Curation

Faculty of the Humanities

University of Cape Town

2025

Acknowledgements

Thank you to my supervisor Michelle Kahn for helping me take the idiosyncratic idea I had and turn it into a workable research project.

Thank you to my partner Shevaun Apps for her love and support and for encouraging me to return to university (again).

In Loving Memory of My Father, Keith Stephen Porter (1942 – 2020).

Abstract

Digital curation requires taking appropriate actions throughout the lifecycle of digital data. Digital curation of research data in a lifecycle is referred to as the custodial approach, while the pragmatic approach is the curation of digital content in everyday life, such as on websites or social media platforms (Dallas, 2016). Online genealogy, in which a genealogist is using a digital platform to construct a family tree and search for source documents, is an example of the pragmatic approach to digital curation. In this study I investigate genealogy through the lens of digital curation to discover how users curate their genealogical information on a major genealogy website, Ancestry.com. In this way, I consider how online genealogy can be studied as an example of the pragmatic approach to digital curation.

This is a qualitative study making use of auto-hermeneutics – I collected data by systematic self-observation (SSO) and analysed it by interpretative phenomenological analysis (IPA). I reconstructed my own family tree in Ancestry.com, evaluating the research hints offered to me by the website algorithm and the member-constructed family trees which contained my ancestors. I found, among other things, that Ancestry.com makes the tree construction process easy and the research hints provided by the website can be verified if caution is exercised, but the same cannot be said for the family trees constructed by Ancestry.com members, many of whom do not practise caution in this endeavour. Study recommendations include a change in the data fields Ancestry.com uses for data capture from members or a change in which it processes and displays that data to ensure that omission of data does not cause incongruous errors to be displayed; and better visibility of the website's guidance on good research techniques to avoid the perpetuation of research errors. Ancestry.com already has verification functionalities which, if applied more widely, have the potential to improve the curation of member-constructed trees.

Table of Contents

List of Figures

List of Abbreviations and Acronyms

BCG	Board for Certification of Genealogists
DCC	Digital Curation Centre
GEDCOM	Genealogical Data Communication
GPS	Genealogical Proof Standard
Gramps	Genealogical Research and Analysis Management Programming System
IPA	Interpretative Phenomenological Analysis
IR	Institutional Repository
LDS	Church of Jesus Christ of Latter-day Saints
mtDNA	Mitochondrial DNA
OA	Open Access
SPARC	Scholarly Publishing and Academic Resources Coalition
SSO	Systematic Self-Observation
Y-DNA	Y-chromosomal DNA

Chapter 1: Introduction

1.1 Introduction to the Study

Genealogy, the study of family histories and lineages, is a popular pastime that has become much easier to practise with the digitisation of records and the availability of these records on internet sites (Hoeve, 2018: 216). Some of these sites allow users to construct their family trees within the site while simultaneously linking digital records, such as birth records or census returns, to individuals within that tree, going so far as to provide suggestions of records which the website algorithm thinks are relevant to the individual (such as Ancestry, 2023). Users seduced by the marketing of these sites, which suggests that genealogy is easy, may build their family trees and histories without critically evaluating the suggestions they are given. This study used digital curation as a critical lens through which to view the construction of a family tree on a major genealogy website.

1.2 Background to the Study

The digitisation of genealogical documents and their availability on the internet has increased access to the documents and made it easier for people to study their family history without the need to travel to the locations housing the physical documents of interest (Thomas, 2007: 2). While the availability of digital records on genealogical websites certainly makes it easy for an inexperienced genealogist to discover potentially relevant information about their lineage, accuracy requires a cautious approach and application of best practices (Knight, 2024: 15). A guidebook for beginning online genealogical research (Thomas, 2007: 27), for example, stresses the importance of research plans, organisational skills, document preservation and citing sources. These skills and activities are also important to scholarly research, and indeed, genealogy has been studied academically, though to a limited degree. It does not, however, constitute an academic discipline in its own right; genealogical research is usually conducted in the context of other disciplines as a multi- or transdisciplinary activity. Despite some presence in academia, there is much that is known in the genealogical community of practice

that has not been extensively studied academically. One example of this is the phenomenon of genealogists constructing their trees online. Genealogy websites have made online genealogical research easier. Inexperienced genealogists may thus overlook the need to verify information, especially when they find this information via the family trees of other genealogists on the website.

1.3 Terminology

Genealogy, or family history, is the pursuit of recording family pedigrees by extracting information from documents, conducting interviews of family members and, more recently, performing genetic ancestry analysis (Durie, 2017: 1; Hoeve, 2018: 215). 'Genealogy' and 'family history' are frequently used interchangeably. Durie (2017: 2) differentiates between genealogy as the basic activity of compiling a family tree using names, dates, places and events, and family history as genealogy understood in its various historical and social contexts and with affective engagement. Alternatively, Yakel (2004) regards the research process as genealogy and the resultant narrative, as retold by the family historian, as family history.

De Groot (2015:104), despite using the terms interchangeably, notes that 'genealogy' is the preferred term in the United States, and 'family history' in Europe. He notes that the usage of 'genealogy' in the United States is related to a perception of the word having professional connotations and 'family history' correspondingly amateur (De Groot, 2015:104). Yakel (2004) notes the same perception, while acknowledging that the terms are used interchangeably, but also inconsistently. Mitchell and Kim (2024: 3), in addition to using the terms 'genealogy' and 'family history' interchangeably, also use the term 'family genealogy'. Darby and Clough (2013: 73) use 'family history research' as a catch-all term, while prefacing the term with 'professional' or 'amateur'/'hobbyist' to be more specific when required. Likewise, Moore (2023: 1) who uses 'genealogy' and 'family history' interchangeably, also prefaces the terms with 'professional' or 'hobbyist'. Hatton (2021: 1) contrasts 'amateur' and 'serious' genealogists – the latter term encompassing all experienced and/or well-trained practitioners, regardless of whether or not they are professional.

In this study I have taken the position of 'genealogy' and 'family history' having overlapping nuances of meaning akin to both Durie's (2017: 2) and Yakel's (2004) definition of the terms, neither interchangeable nor entirely distinguishable. I have mostly defaulted to 'genealogy' and 'genealogist' and used 'family history' and 'family historian' to imply a broader meaning, such as a narrative that includes, but goes beyond, the family tree.

1.4 Digital Curation

Yakel (2007: 335) defines digital curation as "the active involvement of information professionals in the management, including the preservation, of digital data for future use." The purpose of such actions is to add value to the data and mitigate against digital obsolescence, while also enabling reuse of data (Beagrie, 2008: 4).

In the late 20th century, the increasing amount of analogue research data being digitised and new data that originated in a digital format necessitated a greater emphasis on the management and preservation of digital data (Sabharwal, 2015: 11; Yakel, 2007: 336). Key to this endeavour is the recognition that curation of digital data has a lifecycle and that digital curation requires taking appropriate actions throughout the lifecycle. The Digital Curation Centre (DCC) created the Digital Curation Lifecycle Model (Higgins, 2008: 134) as a means to train digital curators and encourage the development of curation methodologies.

Sabharwal (2015: 11) and Dallas (2016: 423) both note that 'digital curation' requires clarification in some contexts, as diverse disciplines attach their own nuances of meaning to the term. This ambiguity extends beyond a scholarly use of the term. Sabharwal (2015: 11) refers to social curation – a form of digital curation that applies to social media platforms – as an example of phenomena to be included in a broader concept of digital curation. Similarly, Dallas (2016: 443) refers to digital curation of research data in a lifecycle as the custodial approach, and contrasts it with the pragmatic approach in which people are curating digital content in their everyday lives on websites, social media platform and cloud services. Thus online genealogy, in which a genealogist is using a digital platform to construct a family tree and search for source documents, is an example of the pragmatic approach to digital curation.

1.5 Research Problem

Professional and experienced amateur genealogists are well aware of the need to critically evaluate each piece of evidence and only ingest new knowledge into their family trees and histories when a weight of evidence is available (Durie, 2017: 4; Hatton, 2021: 13). Genealogy websites attempt to make this process easier by utilising algorithms to link individuals on users' family trees to relevant genealogical records. While such linkage suggestions can be a powerful tool, without careful evaluation these may cause unverified information to be incorporated into family histories.

An increasing number of people are relying on the genealogy industry to help them answer questions of identity (Bottero, 2015: 540). It follows that they should not arrive at incorrect conclusions about themselves as a result of unverified information. That this occurs is well known in the genealogy community of practice (Knight, 2024: 13), but not a matter of academic research interest. The problem that this study will address then is the phenomenon of inexperienced or incautious genealogists incorporating incorrect information into their online family trees without verifying information they receive from website algorithms.

1.6 Study Objective

The study objective was to investigate the practice of genealogy through the lens of digital curation, thus critically investigating the study problem.

In this study I approach this problem by investigating the construction of family trees on Ancestry.com. Ancestry was launched in 1996 (Creet, 2020a: 74) and is today the largest for-profit genealogy company, with a collection of 60 billion records and a DNA genetic ancestry network of 25 million people (Goldman Sachs, 2024).

1.7 Research Questions

The study responded to its research objective by analysing how a major genealogy website (Ancestry.com) offers research hints to its users and also how those users (myself included) curate their genealogical information on the website.

In this study, I ask the following research questions:

RQ1: What is the ease with which a user can produce a family tree on Ancestry.com?

RQ2: To what extent are website-generated suggestions for the family tree correct or to what extent can they be verified?

RQ3: To what extent can users verify the information produced by other users for use in their own family trees?

RQ4: What suggestions can this study offer for the digital curation of these records which would result in a more accurate portrayal of a family tree?

1.8 Significance of the Study

The routes by which erroneous information is ingested into family trees are well known in the genealogy community of practice (Knight, 2024: 13). However, due to the perception of genealogy as a leisure activity or amateur pursuit, it has not been a matter of academic research interest in its own right (Hershkovitz, 2012: 6). Thus, transdisciplinary research with a genealogical aspect to it is typically conducted under the auspices of the recognised academic discipline it is partnered with. The availability of digital records on genealogical websites and the introduction of direct-to-consumer genetic ancestry testing products have broadened the popularity of genealogy as an activity (Hjorthén, 2022: 4). As genealogical discoveries may impact personal identity and relationships with family members and cause distressing emotional responses (Mitchell and Kim, 2024: 6), it is important that anyone researching their family history not be misled by erroneous information. I aim to raise

genealogy's academic profile by highlighting an avenue of research presented by the digital curation aspects of online genealogy. This study will contribute to the academic literature on both digital curation and genealogy

1.9 Overview of Research Methodology

I employed an interpretive research paradigm, as I sought to interpret and understand the genealogical research process, and not explain or predict aspects of it (Babbie and Mouton, 2001: 33). Using a qualitative approach, I observed instances of what could happen in the course of a genealogist constructing their family tree online, and placed such instances in context.

My research method is auto-hermeneutics, described by Gorichanaz (2017: 3) as a synergy of systematic self-observation (SSO) (Rodriguez and Ryave, 2002) and interpretive phenomenological analysis (IPA) (citing Smith, Flowers and Larkin, 2009). SSO is a data collection method concerned with meticulous observation of one's own behaviour, which I employed by noting my observations and reactions as I constructed my family tree on Ancestry.com. IPA is an inductive and iterative research approach with a philosophical grounding in phenomenology, especially hermeneutic phenomenology (Tuffour, 2017: 3), which I used by engaging with my data through multiple iterations to describe the common themes underlying my observations and reactions.

1.10 Delimitations of the Study

As I am using a first-person methodology, my research population was ostensibly myself. My site of study was Ancestry.com, in which I constructed my family tree. My lens was digital curation, in that I studied the curation aspects of online family tree construction.

1.11 Structure of the Report

This study consists of five chapters as outlined below.

Chapter 1: Introduction. The background of the study is introduced, with respect to genealogy and digital curation. The research problems and questions are described. The significance of the study is explained and the research methodology is summarised.

Chapter 2: Literature Review. A review of previous studies is discussed, with respect to genealogy and digital curation.

Chapter 3: Research Methodology. The research approach, design and methods are described.

Chapter 4: Results. Research results are presented thematically.

Chapter 5: Discussion. Research results are discussed in the context of previous studies. Conclusions and limitations of the study are presented and suggestions are made for future studies.

1.12 Chapter Summary

In this chapter, I presented a background to the study, with respect to genealogy and digital curation. I described the research problem and questions. I summarised the research methodology and explained the significance and delimitations of the study.

Chapter 2: Literature Review

2.1 Introduction

Reviewing the available literature around a subject allows the researcher to determine the scope of work that has been done, the patterns that link items in the literature together, and identification of areas where the literature is silent or lacking (Booth et al., 2022: 2).

To uncover scholarly research concerning genealogy on scholarly databases, I conducted literature searches for 'genealogy', 'family AND history' and 'family AND tree'. As 'genealogy' can also refer to a philosophical concept, I also searched for 'family AND genealogy' and 'genealogy AND pedigree'. I conducted narrower searches too, such as 'ethics AND genealogy'. As there is much genealogical information that is derived from genealogy's community of practice, I consulted the websites of prominent genealogy service providers and the blogs of prominent genealogists.

For scholarly research concerning digital curation, I used the search strings 'digital curation' and 'research data management'. I also conducted searches such as 'genealogy AND curation' to find literature that explored the interface of the two subjects.

Resultant from the above searches, this chapter covers important concepts in genealogy, including its history. I also explore genealogy's coverage, or lack thereof, in academic research and explain how it can be investigated from the perspective of digital curation.

2.2 History and Development of Genealogy

Thomas (2007: 1) describes how genealogical research used to require extensive travel to individual archives that housed documents of interest and that each visit involved writing in notebooks and having to work through collections that had not been indexed. In this section I review some of the changes in genealogy that have broadened its reach. Before that, I begin with an acknowledgement.

2.2.1 Acknowledgement of Western Bias

The recording of human pedigrees, and the importance attached to them, is an ancient practice, as judged by the genealogies in *The Bible* (Zimmer, 2018: 163) and the records kept of the descendants of the philosopher Confucius (Creet, 2020b: 26). Zimmer (2018: 472) contends that the origin of genealogy lies in the use of ancestry to legitimise leaders' claims to power.

It is important to acknowledge here that many cultures have rich oral traditions of their history and origins, a key part of which is genealogy. The Māori, for instance, have *whakapapa*, a genealogical concept that can be applied to everyone and everything and is the basis of their knowledge system (Mahuika, 2019: 1). Traditional Chinese genealogies recorded and compiled by families are known as *zupu* (Shiue, 2016: 465) and follow the male family lineage. Similarly, in Korea, *jokbo* are official genealogy books written to record the details of elite families (Lee and Park, 2019: 31). In Africa, traditional oral poets, such the *iimbongi* of Nguni-speaking Southern African groups and *griots* of West Africa, use their art form to praise current or past leaders or heroes, incorporating their (the leader's or hero's) genealogy in the performance (Kaschula, 1999: 56).

In Western Europe, starting in medieval times, the keeping of genealogies ensured that the inheritance of wealth, titles and social standing was maintained (Gomes, et al., 2021: 5). Although genealogy is both ancient and ubiquitous, the origin and epistemology of 'modern' genealogy, with its emphasis on written primary sources of information are "profoundly Western" (De Groot, 2015: 106). While I approach genealogy here in that sense to link it to digital curation, I also acknowledge that, as someone of primarily English descent, it is mostly the records of English origin that have allowed me to trace my own ancestry.

2.2.2 Democratisation of Genealogy

Increased access to records and the availability of new online tools with which to study one's ancestors has opened up genealogy to the public at large, where once it was the preserve of

an elite few (Durie, 2017: 3). This is popularly described as a 'democratisation' of genealogy (Hjorthén, 2022:2). In the broadest sense, the democratisation process started when governments began systematically to keep records of their subjects. Thus the United Kingdom and the United States of America both began collecting detailed census information in the mid-nineteenth century. Birth, marriage and death records, previously kept mainly by churches (along with baptism and burial records), became a bureaucratic preoccupation of governments around the same time (Chater, 2008: 30). Likewise, changes in society and the law resulted in a more systematic and centralised keeping of criminal, legal, migration and military records.

In 1830, around the same time the changes described above were happening, Joseph Smith formed what would become known as the Church of Jesus Christ of Latter-day Saints (LDS) (Creet, 2020c: 40). Early in their history, the LDS adopted the practice of baptism for the deceased: a baptised Latter-day Saint can later be baptised in a LDS temple on behalf of a deceased family member (Chidester, 2000: 444). A consequence of this doctrinal practice was the collection and preservation of genealogical records. Indeed, it is a LDS requirement that Latter-day Saints actively research their genealogies to enable such baptisms. As a result of these endeavours, the LDS's FamilySearch Library (previously known as the Family History Library) in Salt Lake City, Utah is the world's largest genealogical collection (Otterstrom, Bunker and Farnsworth, 2021:4) and houses "… 1.3 million rolls of microfilmed genealogical records; 190000 microfiche; 340000 books, serials and other formats; and 125000 periodicals…" as well as "… the names of more than 8 billion deceased people from more than 100 countries" (Intellectual Reserve, 2024a). The five-floor facility, covering over 13000 m^2, offers free access to the public.

2.2.3 Digitisation and Commercialisation

The next chapter in the democratisation of genealogy is the digitisation, online availability and commercialisation of collections. Having preserved any records to microfilm since 1938 (Hoeve, 2018: 215), the LDS began digitising them in 1998 and launched the website FamilySearch.org in 1999, through which it provides access to these records free of charge.

As of 30 December 2024, the website offered free access to more than 20.5 billion records and images (Intellectual Reserve, 2024b).

Ancestry.com came online in 1996, several years after its founders (LDS members Dan Taggart and Paul B. Allen) started selling floppy discs containing LDS genealogical material, which grew to a suite of products including software, databases and reference materials (Creet, 2020a: 74). Once they realised that the World Wide Web was the next logical location for these sorts of materials, they launched the Ancestry.com website. It grew quickly and Ancestry (the company) expanded its reach by acquiring other genealogical websites (Creet, 2020a: 83). Ancestry is today the largest for-profit genealogy company, with a collection of 60 billion records and a DNA genetic ancestry network of 25 million people (Goldman Sachs, 2024).

Further democratising of genealogy thanks to digitisation was the creation of other websites. Two major commercial websites are FindMyPast.com (Findmypast, 2025) and MyHeritage.com (MyHeritage Ltd, 2025), established in 2006 and 2003, respectively. Each of these websites offer an extensive records database and family tree building functionality. Ancestry.com and MyHeritage.com integrate those functionalities with their respective genetic ancestry testing services, which FindMyPast.com does not offer.

Ancestry.com acquired RootsWeb.com in 2000 (Creet, 2020a: 81). Established in 1996 and developing from message boards and mailing lists of the nascent internet, RootsWeb.com was considered a pioneer of internet genealogy. The site functioned as a means for genealogists researching the same families to communicate and collaborate and all data on the site was uploaded by the users themselves (Starasta, 2009:54). Not only did Ancestry.com not invest in the maintenance of RootsWeb.com, it also began to remove features from the website (O'Brien, 2020) and halted uploads of new information to RootsWeb.com. When Ancestry.com announced that the RootsWeb.com mailing lists were to be shut down on 6 April 2023 and the family trees on the site removed on 15 April 2023, Estes (2023) feared that Ancestry.com would delete these components, or RootsWeb.com in its entirely, causing a massive loss of genealogical information. However the RootsWeb.com trees were later migrated to Ancestry.com, meaning that despite their temporary absence from the internet, those trees were not permanently lost as resources to genealogists.

In addition to the large commercial genealogy sites that have been mentioned and which receive much attention, there are a multitude of online resources on genealogy available. Many of these resources provide access to information not available from the larger sites or the LDS collection. South African examples are Ancestors.co.za (Ancestors Research South Africa, 2019), the professional website of genealogist Heather MacAlister, who hosts a database containing many South African collections that she has been responsible for researching and digitising, and 1820settlers.com (Tanner-Tremaine, n.d.), a site tracing the descendants of the British settlers arriving in South Africa in 1820. Websites such as these are important, as their local focus provides coverage of resources that might not be available via the larger companies.

2.2.4 Genetic Ancestry Testing

The final component of democratisation of genealogy is the introduction of direct-to-consumer genetic ancestry testing (Leonard, Macdonald and Holton, 2019: 1), provided by, amongst others, Ancestry.com and MyHeritage.com. These tests allow people to have their DNA analysed and compared to the DNA of others. The use of these results (genetic genealogy) to complement traditional family history research has allowed genealogists to break down so-called 'brick walls' – a line of research that has come to a dead end, such as a lineage that cannot be traced any further or an ancestor for whom there is little information and searches reveal nothing further (Thomas et., 2007: 29).

Direct-to-consumer genetic ancestry testing makes use of saliva samples or buccal swabs, from which DNA is isolated. Data is then produced and analysed from four types of DNA in human cells (Leonard, 2019a: 32). The nucleus of the cell contains 23 pairs of chromosomes. The twenty-third pair are the sex chromosomes: XX or XY. The other 22 chromosomes are termed autosomes. Y-chromosomal DNA (Y-DNA) is male-specific and provides a marker of paternal descent. It was the first DNA type to gain widespread use due its applicability to patrilineal surname research (Macdonald and Cleary, 2019: 104). Autosomal DNA testing is by far the most common, data rich and generally informative (Leonard, 2019b: 60), allowing

an estimation of familial relationship to other users, termed ‘DNA matches’, with whom they share segments of DNA inherited from a common ancestor.

Another popular use for autosomal DNA tests is to obtain a percentage breakdown of geographic ancestral origins (Ancestry, 2024b), marketed and described in terms of ethnicity. X-chromosomal DNA is not tested in isolation, but is generally included in autosomal tests. In human cells, but outside of the nucleus, are cellular components called mitochondria, which also contain DNA. Mitochondrial DNA (mtDNA) is a marker of maternal descent, although it is present in all people (Leonard, 2019a: 32).

In addition to Ancestry.com and MyHeritage.com, the other major genetic ancestry testing companies are 23andMe (23andMe Inc., 2025), FamilyTreeDNA (Gene by Gene Ltd, 2025) and LivingDNA (Living DNA Ltd, n.d.) While most companies focus on autosomal tests, FamilyTreeDNA also provides detailed Y-DNA and mtDNA tests. The websites of these companies, with the exception of Ancestry.com and 23andMe.com, also accept uploads of the raw DNA data provided by other companies, so that customers can make use of their tools and connect to a broader base of DNA matches. All of these websites offer suites of features that allow their members to further analyse their DNA and their relationship to their DNA matches. One notable feature common to most sites, but not Ancestry.com, is a chromosome browser, which shows graphically where users share DNA segments inherited from a common ancestor. GEDmatch.com (GEDmatch, 2024) is not a DNA testing service itself, but accepts uploads of both raw DNA data and family tree files. It also hosts a variety of third party genetic analysis features.

I have described important developments in genealogy to place the current study in context and provide background to the state of genealogy at the present time. The genealogist of today has powerful tools available to them to aid them in their research.

2.3 Digital Curation

The concept of digital curation grew out of the collaborative efforts of scientists and librarians who recognised the need to preserve and manage research data over the long-term (Beagrie, 2008: 4), culminating in a pivotal 2001 conference where specialists in the emerging field decided on the term 'digital curation' to describe the actions necessary to ensure the long-term usability of digital data. In addition to her definition, Yakel (2007: 338) describes digital curation as "the umbrella term for digital preservation, data curation and digital asset and electronic records management."

2.3.1 Philosophy of Digital Curation

Reviewing early definitions of digital curation, Yakel (2007: 338) identified five important concepts and activities:

- The curation of digital material has a lifecycle. Poole (2016: 964) noted eight lifecycle models worthy of merit, the most prominent of which was the DCC Curation Lifecycle Model described by Higgins (2008: 136).
- Creators of digital records should remain actively involved in their curation.
- Digital material needs to be periodically appraised to determine if it is still worthy of preservation.
- Long-term access to data must be ensured.
- Digital material must be preserved so that they remain accessible and usable.

I discuss the DCC Curation Lifecycle Model in a later section. Here I highlight that the points above emphasise the commitment of digital curation, as a discipline, to sharing of data and open access (OA) publication of research. OA assists in more rigorous and efficient research and helps research disseminate faster and have more impact (Poole 2016: 106). Data that is openly accessible allows for its interrogation and reuse and improves the reproducibility and reliability of research (Poole, 2016: 107).

2.3.2 Institutional Repositories and Open Access

The need for improved digital infrastructure also served to promote the establishment of institutional repositories (IRs) – digital archives through which research institutions take control of their digital output (most notably publications and research) by taking responsibility for its preservation, management and dissemination (Lynch, 2003: 328). Lynch (2003: 327) singles out the Massachusetts Institute of Technology's DSpace as a pioneering IR that encouraged other institutions to emulate its model by publicly releasing the open source IR software that was used. Crow (2002:4), writing on behalf of The Scholarly Publishing and Academic Resources Coalition (SPARC) in a position paper, advocated for IR as a means for research institutions to showcase the quality of their research, adding to their prestige. Lynch (2003: 330) argues that an IR relieves researchers of an administrative burden by handing over the preservation and management of their research output to information specialists.

When Graham (1995: 331) advocated for what he termed 'digital research libraries', he envisioned IRs as a tool to reform scholarly publishing and mitigate against the effects of the 'serials crisis'. At the time of writing his article, academic libraries had begun to feel the effects of price increases of scholarly journals, leading many academic libraries to cancel some of their journal subscriptions. The advent of the internet had advanced the idea that electronic journals could result in academic institutions taking back control of the publication of their research from commercial publishers (Yiotis, 2005: 158). Discussions around this topic led to the establishment in 1998 of the aforementioned SPARC to promote IRs and assist journals in adopting an OA publishing model; and in 1999 of the Open Archives Initiative, which developed standards for interoperability between IR providing OA archive content (Yiotis, 2005: 159).

2.3.3 Research Data Lifecycle

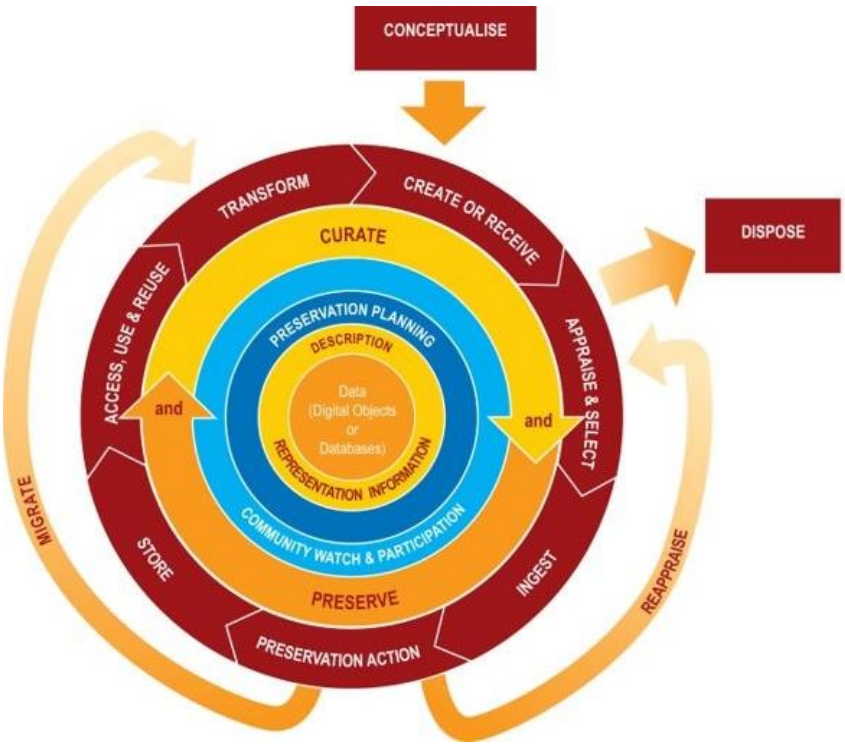

Figure 2.1. The DCC Curation Lifecycle Model (Higgins, 2008: 136)

The DCC developed their Curation Lifecycle Model (shown in Figure 2.1; Higgins, 2008: 136) to represent the stages in the curation of digital material and to provide a planning tool for successful curation. Data, in the form of databases or digital objects, are placed in the centre of the lifecycle model (Higgins, 2008: 137). Digital objects can be simple, such as a single file with its associated metadata, or complex – comprising a combination of digital objects, such as websites.

Metadata is structured data that describes a data object (Mayernik, 2020: 699). The concept of metadata has its origins in the description of catalogued items, most notably of books in libraries (Haynes, 2018: 4). In the context of digital curation, it refers to information encoded into the digital object that facilitates use of that object. Complete description of digital material with appropriate metadata is critical for successful curation (Higgins, 2012: 38). The nature of the object and the metadata standard used will determine what metadata is captured.

The types of actions applied to data are represented as concentric circles in the lifecycle model. A number of actions are performed sequentially (Higgins, 2008: 137):

- Conceptualise: This is the initial planning stage in which decisions are made around how data is created, captured and stored. Sabharwhal (2015: 97) describes the three components of conceptualisation as the people involved, the content to be curated and the technology required. It is at this stage that issues around intellectual property, copyright and interoperability must be addressed (Higgins, 2012:23).
- Create or Receive: This is the creation of new data or the acquisition of data from other sources, such as an archive (Higgins, 2008: 138). For new data, technology specifications will determine the file formats and metadata standards used (Higgins, 2012: 25; Sabharwal, 2015: 104). Collection development policies must be in place to determine what data from other sources can be deposited (Higgins, 2012: 26).
- Appraise and Select: A determination is made by the researcher whether the data in question holds any long-term value and should be retained in an archive or repository (Higgins, 2008: 138; Sabharwal, 2015: 105). Decisions can be subjective, but an appraisal policy should be in place to keep collections relevant to their user community (Higgins, 2012: 29).
- Ingest: This refers to the formal accession of selected data from its creator or collector to an archive or repository and subsequent indexing to make the data searchable (Higgins, 2008: 138; Sabharwal, 2015: 105). Identification of who is responsible for curation tasks must be determined at accession (Higgins, 2012:30). Where collections are donated or transferred, laws governing the transfer of intellectual property will apply (Sabharwal, 2015: 105).
- Preservation Action: Since the aim of digital curation is to ensure long-term preservation of data of value and ensure that it remains authoritative, this stage is an important one (Higgins, 2008: 138). Actions may include migrating data to new file formats or use of emulation software on a new platform to enable preservation (Higgins, 2012: 32).
- Store: Data must be stored securely according to relevant standards (Higgins, 2012: 35). The nature of the data storage depends on the resources of the institution – if

storage is in a repository, the choice of operating systems and content management systems must also be taken into account (Sabharwal, 2015: 105).

- Access, Use and Reuse: This stage covers matters of daily accessibility, including access control and authentication procedures (Higgins, 2008: 138). Redaction or anonymisation of sensitive or personal information may be required in some instances (Higgins, 2012: 37). In order for data to be discovered, appropriate search tools should be in place.
- Transform: Derived data may be created from the original, either through migration to another file format or by creation of a subset of the data to consider separately (Higgins, 2008: 138). Technologies advance, along with revision of file formats and the software applications in which they operate, making timely file format migration necessary to prevent incompatibility of file formats (Sabharwal, 2015: 107).

Certain lifecycle actions are not performed sequentially, but throughout the lifecycle (Higgins, 2008: 137):

- Description and Representation Information: The assignment of metadata using applicable standards ensures that all data is properly described and can be appropriately controlled. Metadata standards support interoperability, while incomplete metadata may cause data to become unusable or undiscoverable (Higgins, 2012: 38). Sabharwal (2015: 101) describes this process as a collaboration between curators and cataloguers and also between the curation department and the academic department that created the data. This collaborative effort must take into account that different collections and departments may use or require different metadata profiles or schemas and may need to include additional documentation to provide all the information needed for effective curation.
- Preservation Planning: All actions in the lifecycle model require plans for their management and administration (Higgins, 2008: 137). Preservation plans should be reviewed and revised annually to keep pace with technology (Higgins, 2012: 32).
- Community Watch and Participation: It is incumbent on creators and managers of digital data to be aware of new developments in their field and cultivate opportunities

for collaboration (Higgins, 2012: 40). Curation also involves stakeholders with an interest in the process whose feedback is valuable (Sabharwal, 2015: 102).

- Curate and Preserve: Sabharwal (2015: 99) laments the omission of conservation as a full lifecycle action and discusses 'Preservation and Conservation' in the place of 'Curate and Preserve'. His focus in this context is the extension of the life of digital content and the migration of content from obsolete media formats. Higgins (2012: 32) omits this as a discrete action.

Three lifecycle actions are described as occasional (Higgins, 2008: 138):

- Reappraise: This refers to the identification of data that should not be preserved in the long-term.
- Dispose: Material and associated metadata might be removed from the archive to be transferred to another archive, or destroyed if it is not considered important enough to store and preserve in the long term. If the data is to be destroyed, this must be done securely, following legal and policy requirements (Higgins, 2008: 138).
- Migrate: Higgins (2008: 138) discusses this action in the context of migrating data to a different format as a preservation action, while Sabharwal's (2015: 108) discussion of this action refers to migration of records or collections to new repositories – accompanied by changes to metadata, migration to new file formats and reorganisation of collections to suit the new repository.

I have described some of the principles and concepts behind digital curation. I shall now draw parallels between it and genealogy and discuss how genealogy can be approached through the lens of digital curation.

2.4 Curating Genealogy

The present study approaches genealogy through the lens of digital curation. Dallas (2016: 442) refers to digital curation in the form that I discussed above – applied to research data, curated by information professionals, deposited in institutional repositories and occurring in a lifecycle – as the "custodial approach". He compares this approach to a pragmatic approach for which he advocates, which refers more broadly to people curating digital material in their professions, communities of practice and personal lives, rather than purely for research. A pragmatic approach to curation takes places on websites, social media and cloud services, and the material under curation is constantly changing and being enriched (Dallas, 2016: 439).

Roued-Cunliffe (2017: 109) includes genealogists as examples of heritage amateurs who build collections of cultural heritage, noting that in their research "they often collect and manage in some form of digital repository" (Roued-Cunliffe, 2017:112). She also lists names, dates and places as genealogical metadata. Thus genealogists practising their craft on an online platform are performing digital curation in the broader pragmatic meaning of the term.

Online genealogy involves multiple layers of curation. Genealogy websites perform digital curation in the custodial sense, in that they store, preserve and maintain files on their servers, both the collections of documents they make available to website members and the family trees that members create online. Website users, in their turn, curate records relevant to their own families, either by selecting material provided by the websites or by digitising and uploading their own material, such as family photos or documents (which in turn may form part of other users' curation activities). The central focus of this activity is construction of a family tree, accompanied by association of specific people with specific records, as a way of citing specific information. Thus genealogists are applying metadata to their family tree data (Roued-Cunliffe, 2017: 112) and attributing information to a source. Annotation of profiles within a family tree (for example, with additional information and anecdotal stories) goes beyond narrow definitions of genealogy (Durie, 2017: 2; Yakel, 2004) to approach what may be called a family history. Online genealogy can comprise multiple linked data sets in need of organisation – family trees, attached records, DNA matches and others. Some users may participate in a third layer of curation – downloading the information they have obtained in

this way to their personal devices and organising it as they see fit (Roued, Castenbrandt and Revuelta-Eugercios, 2023:200).

Describing a person's position in a family tree sometimes requires a formal numbering system, just as data objects require unique identifiers as part of their metadata. Thomas et al. (2007: 36) describes two such systems. The Ahnentafel or Sosa-Stradonitz system is a traditional one used for listing people in an '*ahnentafel*' – from the German term for 'ancestor table'. The subject of the *ahnentafel* is numbered 1, their father 2, their mother 3 and so on. For any person in the *ahnentafel*, their father's number will be double theirs and their mother's number will be double theirs plus one. In contrast, the Henry system starts with a progenitor and lists their descendants. The progenitor is 1 and their children are described by the addition of an alphanumeric symbol: thus their first child is 11, their second is 12 and so on. 11's children are 111, 112 and so on. If 11 had more than nine children, then the tenth, eleventh and twelfth will be 11X, 11A and 11B respectively.

Curation of a family tree requires a sense of who belongs on that tree. If the purpose of the tree is to determine the relationship to a distant relative, Leonard (2019b: 69) recommends building the tree "wide and deep" to cover all possibilities. Likewise, Johnson Crow (2016) favours an inclusive approach to tree building. Erben (1991: 279) and Durie (2017:3) each express cynicism toward and question the motives of genealogists with large trees. There is also an ethical issue of who one should include on a tree and how they are represented (Moore, 2023: 6; Knight, 2024: 5).

Digital curation of online genealogy also applies to the file types used. The open text-based file format for the building of family trees called GEDCOM (which stand for Genealogical Data Communication), with the file extension .ged (Intellectual Reserve, 2021), is now a de facto standard in online genealogy and allows easy interoperability between various online platforms and software packages. An alternative format (Gramps XML; file extension .gramps) was created for Gramps (standing for Genealogical Research and Analysis Management Programming System), a free and open source software package (Gramps, 2022).

In summary, genealogy has been linked implicitly to digital curation by the language used to describe genealogical processes and concepts (Roued-Cunliffe, 2017: 112) and by the examination of what constitutes digital curation, such that online genealogy falls into a broad category of 'pragmatic' digital curation. True to this association, the practice of genealogy utilises formal numbering systems and specific file types.

2.5 Genealogical Research

In this section, I discuss genealogy as a research topic and review studies that have been conducted under its umbrella. Academically, genealogy is rarely considered as a discipline in its own right, being multidisciplinary or transdisciplinary by nature. Hershkovitz (2011: 22) points out that researching one's ancestor or family may be of interest to the individual researcher, but will not be academically interesting unless it utilises a novel methodology or has wider implications and avenues for future research. Thus genealogical research is invariably conducted in the context or under the umbrella of another discipline, with its research approaches varying accordingly (Roued, Castenbrandt and Revuelta-Eugercios, 2023: 189). Roued and Klareld (2025: 248) observe that prior to reviewing the research conducted on family historians, they believed that little work of such a nature had been published. However, their review of the literature demonstrated that while more abundant than believed, research on family historians is "heterogeneous and disparate" due to genealogy's multidisciplinary nature.

2.5.1 Genealogy Practice and Research Behaviour

New academic journal titles dedicated to the study of genealogy have been introduced, such as *Genealogy*, for which Durie (2017) is an introductory article. While acknowledging that genealogy is progressing towards academic recognition, he also highlights the scarcity of dedicated university departments and the lack of professional qualifications beyond certifications of competence (Durie, 2017: 6). As the practice of genealogy happens largely

outside of academia, there is much that is known within the genealogy community of practice that has not been formally researched. This knowledge disseminates through the community in the form of genealogy magazines, online discussion groups, blogs, YouTube vlogs and conferences, the most prominent of which is the annual RootsTech conference held in Salt Lake City, Utah (Intellectual Reserve, 2024c). These outlets help to promote good practice.

The Board for Certification of Genealogists (BCG) is an American genealogical corporation which establishes standards of competence and ethics for genealogists and certifies genealogists who meet those standards (Board for Certification of Genealogists [BCG], 2023a) The BCG established the Genealogical Proof Standard (GPS) as a set of five principles a genealogist must follow in order for their work to be credible (BCG, 2023b), as follows:

- First, research must be reasonably exhaustive, having examined as many sources as possible.
- Second, all information items must have complete and accurate source citations.
- Third, all evidence must have been analysed, correlated and interpreted.
- Fourth, all conflicting evidence must have been resolved.
- Five, a soundly reasoned and coherently written conclusion must be drawn.

Ancestry.com (2024c) provides a guide to accurate research, which it has adapted from these principles, in their online support centre. Some experienced and/or diligent family historians pride themselves on applying high standards in their research to raise confidence in their results (Willever-Farr and Forte, 2014: 481), though it is not clear whether they are explicitly familiar with the principles of the GPS. One guidebook for practising genealogy (Thomas et al., 2007: 26) explicitly sets out a 'Family History Research Cycle' to describe the research process:

- Planning: decide which part of the family tree to research, write down what you already know and determine the resources you will need to find out more.
- Collecting: looking for photographs and documents and interviewing family members to find out what they know.
- Researching: searching for information online and in libraries and archives.
- Consolidating: incorporating new information into a central location.

- Distilling: take stock of the information gathered and discard what does not belong.

Roued and Klareld (2025) categorised four approaches to research about family historians:

- Literature from the mid-to-late 20th century consisted of studies of family historians as users of libraries and archives, with an emphasis on their access to resources. This is the historical approach.
- The practice approach studies how family historians go about their research, sometimes constructing models of the research process. This approach differs from the historical approach in that new technologies and developments, such as home computing, the internet and social media, commercialisation of archives and DNA testing have changed the nature of genealogy.
- Genealogy is a personal endeavour in which discoveries, both welcome and unwelcome, can arouse strong emotions. The affective approach studies this aspect of genealogy.
- The critical approach studies power relations, sociocultural norms and inherent biases in genealogy, demonstrating how traditions and representations can be problematic.

The following paragraphs look at examples of the practice approach as per Roued and Klareld (2025: 249), as the present study aligns most closely with this approach.

Information-seeking behaviour among genealogists was studied by Duff and Johnson (2003: 94), who used semi-structured interviews to investigate how genealogists use physical archives. They found that, while professional genealogists were adept at finding resources in archives, amateurs devised ways of obtaining information to circumvent their difficulty in using archives, relying on networks of colleagues to help them find information.

Darby and Clough (2013: 75), who used qualitative questionnaires and unstructured interviews, designed a model of the phases genealogists go through during the course of their research process. They found that, after collecting family photographs and documents to find information to start their genealogical search, genealogists learned the process of doing

research; they started with easy tasks and easily accessible resources, progressing in phases to more challenging research tasks as they learnt more about the research process and the resources they needed to continue their research. The final phase that in which they conduct difficult research to push back particular ancestral lineages. Genealogists revert to earlier phases of the process when new avenues of research open up. The model thus "describes the iterative and cyclical nature" of genealogical research (Darby and Clough, 2013:78).

Friday (2014) conducted a primarily ethnographic study to investigate how genealogists conduct online research. She developed an iterative model of online research behaviour in which the genealogist takes an action, such as locating a source or searching for particular information, using multiple strategies as the research requires, until they reach an outcome (Friday, 2014).

Yakel (2004) and Yakel and Torres (2007) investigated how genealogists manage their research output. Both studies utilised semi-structured interviews, while the latter also included in-person observation of participants performing genealogy tasks, alone and in group meetings. Yakel (2004) found that genealogists see their work as a continuous process, not something that can be 'finished'. For the current study, it is interesting that Yakel (2004) refers to how genealogists "manage research data". Yakel and Torres (2007: 105) described a major activity of genealogists as their search for and collection of records pertaining to their families, including oral accounts, and found that genealogists were eager to help their colleagues do the same.

Willever-Farr, Zach and Forte (2012: 304) studied genealogists' interactions on Ancestry.com message boards to understand how they cooperate in their research. Rather than ask how to find information, genealogists were more likely to ask for (and receive) specific information that they could immediately incorporate into their trees, such as data from census or marriage records (Willever-Farr, Zach and Forte, 2012: 310).

Roued, Castenbrandt and Revuelta-Eugercios (2023: 197) used questionnaires and focus group interviews to examine family historians' utilisation of digital platforms, finding that Danish genealogists make use of a "buffet" of online platforms, using whichever one suits their purposes in their research at that point in time. Roued, Castenbrandt and Revuelta-

Eugercios' (2023) model of the research process begins with an initiating event, followed by the collection of basic information (births, marriages and deaths) and enrichment – the collection of a wider range of life facts, such as criminal records. During and after collecting and enriching, researchers practise information management, such as preserving information sources for future use, and community engagement, including social media interactions.

The studies described here agree that genealogy is a cyclical and iterative process, usually without an endpoint. Collaboration between genealogists is an important feature and genealogists approach their work with much affective engagement.

2.5.2 Genealogical Websites as a Source of Data for Research

The databases of Ancestry.com and other genealogical websites have served as a source of data for a range of research topics. Fukushima, Bourrier and Parker (2022: 11), for example, studying the letters of Victorian novelist Dinah Craik, used census records from Ancestry.com and FindMyPast.com to link names in the letters to household staff living with her. Other genealogical documents on those websites helped to provide more details of the lives of these staff members. Similarly, Brach (2008: 246) used data from Ancestry.com to study the lives of 50 historical Harvard University botanists, including the dates on which they departed or arrived on research travel journeys.

Census data, death certificates and obituaries from Ancestry.com allowed Wortzel et al. (2021: 3) to study historical asbestos exposure in an asbestos manufacturing community, demonstrating how social inequalities in the town of Ambler, Pennsylvania were responsible for some of the patterns of exposure seen. Raeburn et al. (2017: 348) similarly investigated the history of mental healthcare at an Australian "lunatic asylum" (as per the paper and historical usage of the term), with Ancestry.com records helping to provide details of patients at an overcrowded facility.

At a larger scale, data-mining of genealogy websites has generated data on child naming trends, lifespans, fertility, multiple births and human geography. Fire, Chesney and Elovici

(2014: 2) data-mined WikiTree.org, a free genealogy website where users can add their family tree to a single collaborative tree, for this purpose, while Kaplanis et al. (2016: 171) did the same with Geni.com, another genealogy website using a collaborative tree model.

Ancestry itself has been the subject of many studies. De Groot (2020) reviewed the history and evolution of the company, its business practices and its advertising from the perspective of public history and the potential impact of research in that field. Putman and Cole (2020: 211) used constitutive rhetoric as a methodological lens to analyse Ancestry television advertising in the United States. They conducted the study to demonstrate how Ancestry appeals to White Americans' longing for "symbolic ethnicity" – a nostalgic identification with a particular strand of their ancestry (Putman and Cole, 2020: 218).

Kriesberg (2017: 11) interviewed government archivists and employees from private companies, including Ancestry, making use of government archives, with the aim of understanding public-private partnerships in archives. Privacy and consent are important issues in genealogy, with Wallace et al. (2015: 2) investigating whether the consent and notification of third parties, not just DNA test takers, was being addressed in the consent and informational material from Ancestry and other genealogy companies. They found that prior to the availability of genetic ancestry testing, the genealogist would involve their family in their research and share their findings within their family. However, genetic genealogy tends to be more individualistic, with less input from family members (Wallace et al., 2015: 7). They suggest that consent forms accompanying DNA tests include clauses recommending test takers consult their families before proceeding. Blom (2022: 340) submitted her own test samples to four DNA testing companies (23andMe, Ancestry, FamilyTreeDNA and MyHeritage) so that she could analyse the narrative given to her by these companies about her ethnic composition, among other information. As a Swede, she criticised the emphasis in the results on Americans with Swedish ancestry. Ancestry.com also presented information to her as part of a narrative of Swedish culture, which also was skewed toward the culture of Americans with Swedish ancestry.

2.5.3 Genealogy's Role in Identity

Theunissen (2022: 5) conducted semi-structured interviews to study how DNA test results affected the identity of test-takers. Some of the participants had discovered that their parents or grandparents had misattributed parentage or they (the parents or grandparents) were not told that they were adopted Theunissen (2022: 8). The participants sometimes identified greatly with a particular part of their family, only to find they had no genetic link to them, causing a re-thinking of their identity.

Stallard and De Groot (2020: 276) gathered qualitative data from focus groups of family historians. One common theme that emerged in the study was the participants feeling an emotional connection to past generations and a sense of obligation towards them. Guerrini et al. (2022: 486) designed a survey to gather quantitative data, which they statistically analysed to uncover the impact of participants' discovery of new information about themselves. More than half of the participants had changed their behaviour in some way because of the discovery – 8% started a new hobby or began a cultural practice related to their discovery, while 5% sought mental health support.

Sleeter (2020: 2) recounts in an editorial how she started applying critical race and critical feminist theories, which she was already using in her academic career, to her family history, coining the term 'critical family history.' Oliver (2021: 2) describes his family history journey through the lens of being a "Black Queer-identified man", while Still (2020: 2) researched her grandmother's life using Ancestry.com, applying Black Feminist Theory. Saunders (2024: 706) analysed episodes of *Who Do You Think You Are?* – a documentary series in which celebrities' family histories are researched – through the lens of race and post-colonialism.

Autoethnography is a qualitative research method that analyses the personal experience of the researcher in order to understand the cultural experience being researched (Ellis, Adams and Bochner, 2011). Much of critical family history can be considered autoethnographic in nature, including some of the studies listed above, although Sleeter (2020: 2) prefers to separate critical autoethnography from critical family history. Also, autoethnography, while reflective, is not necessarily critical. Gastmeier (2018: 536), as a mixed-race Canadian-Jamaican, returned to Jamaica to investigate her family history and heritage as a way of

exploring her identity. Halley's (2011: 285) autoethnography mixed genealogy with the memory of a childhood trauma in order to make sense of the experience. A more unusual approach is described by Meyer (2020: 2), who studied her family history from a critical perspective, but then wrote her narrative as a novel.

2.6 Chapter Summary

In this chapter I have described the developments in the history of genealogy with respect to how all people, not only powerful elites, have the means to investigate their ancestry. Thereafter I discussed important concepts in digital curation, such as institutional repositories, open access, metadata and the DCC's Curation Lifecycle Model and discussed points of contact between digital curation and genealogy. I described aspects of the practice of genealogy and models of how genealogical research is conducted. I also reviewed studies which investigated how genealogy affects identity and studies which made use of data from Ancestry.com and other genealogical websites.

Chapter 3: Research Methodology

3.1 Introduction

In this chapter I outline my research paradigm and approach and which research design I employed to investigate the study problem. I describe my research methods, including how I collected my research data and subsequently analysed it. I set out the ethical considerations involved and the measures I put in place to enhance research rigour.

3.2 Research Paradigm and Approach

In this study, I approached genealogy through the lens of digital curation, focussing on the curation aspects of a phenomenon (construction of family trees) within a culture (genealogy). I employed an interpretive research paradigm, as I sought to interpret and understand the genealogical research process, and not to explain or predict aspects of it (Babbie and Mouton, 2001: 33). In so doing, I acknowledge that, irrespective of any objective historical reality, the findings of this study are subjective, and emerge from the meanings I attach to them.

I observed instances of what could happen in the course of a genealogist constructing their family tree online, and placed such instances in context, rather than measuring how prevalent such activities were. This is the idiographic research strategy (Babbie and Mouton, 2001: 272), characteristic of the qualitative approach, which is what I employed in this study. Qualitative research, unlike quantitative research, does not seek numerical data to predict phenomena or to determine cause and effect (Tisdell and Merriam, 2016: 5), but rather seeks to understand how people interpret or attach meaning to experiences. Where I included numerical data, it was as an adjunct to descriptive observations, which comprised the bulk of my data.

3.3 Research Design

Of the research designs that work within the interpretive paradigm and qualitative approach, a number could have been applicable here. Phenomenology is the study of how people consciously experience their world (Schurink, Schurink and Fouché, 2021: 293), in which researchers analyse participants' descriptions of their experiences. Of particular interest to me was hermeneutic phenomenology, which asserts that people's preconceptions cannot be separated from how they experience their world (Tuffour, 2017: 4). Ethnography, used in cultural anthropology, makes use of observational data by a researcher embedded in a culture (Babbie and Mouton, 2001: 279). Autoethnography, a form of first-person ethnographic research, emphasises the researcher's personal experience. However, another first-person methodology grounded in hermeneutic phenomenology and partly influenced by autoethnography has been advocated for use in information science and seemed fitting for this study: auto-hermeneutics (Gorichanaz, 2017).

Gorichanaz (2017: 3) described auto-hermeneutics as a synergy of systematic self-observation (Rodriguez and Ryave, 2002) and interpretive phenomenological analysis (citing Smith, Flowers and Larkin, 2009). SSO is a data collection method concerned with meticulous observation of one's own behaviour. IPA is an inductive and iterative research approach with a philosophical grounding in phenomenology, especially hermeneutic phenomenology (Tuffour, 2017: 3). Using IPA, a researcher searches their data for layers of meaning, while trying to get to the heart of the phenomenon they are studying. Gorichanaz (2017: 4) notes that although IPA was designed to be used in conjunction with semi-structured interviews, it is suitable for a variety of data sources.

Genealogy, as popularly practised, is a personal endeavour with close affective engagement with the study material (De Groot, 2015: 104). As someone who has been using online platforms to research their family history for many years, I am able to study the subject from within the community of practice, not merely as an external observer. It is appropriate then to study genealogy, in this case specifically my own genealogy, using a first-person research methodology that shares that affective engagement, hence my adoption of auto-hermeneutics as a research design.

3.4 Research Methods

I collected research data by systematic self-observation (as per Rodriguez and Ryave, 2002) of my reactions to user-produced family trees, user profiles and website-generated suggestions of relevant documents when I created a new family tree in Ancestry.com. An advantage of SSO is that minutiae can be captured in the moment and not lost as they might be if the data collection method was a survey or interview. For some activities, such as ultramarathon running, which is the example provided by Gorichanaz (2017: 5), it is not practical or convenient to capture such minutiae in the moment but rather at an opportune time soon after. In this study I captured my reactions in the moment, while reviewing Ancestry.com research hints or member-constructed family trees.

Rodriguez and Ryave (2002) recognise that study topics for which SSO is a suitable method of data collection have specific characteristics: the phenomenon being studied must be of short duration; have a defined beginning and end; be intermittent in occurrence; be noticeable (to be able to observe it) and "should be a single, focused phenomenon that is natural to the culture." I contend that constructing a family tree on Ancestry.com, in the context of this study, satisfies the five criteria. While genealogy as an activity might not have a short duration or a defined beginning or end, this study did.

3.5 Data Collection

As I used a first person methodology, my research population was ostensibly myself – I collected and analysed my own observations. I created a new family tree of myself and my direct ancestors on Ancestry.com. This tree was populated initially with myself and my parents and grandparents and served as input data from which Ancestry.com suggested connections to members, their family trees and other suggestions (output data). Thereafter, I individually and systematically added my known ancestors from previous generations, observing and noting how dates, names and places were listed on the Ancestry.com research hints offered to me. In particular, I noted how these items of information differed between sources pertaining to the same person. I also observed how Ancestry.com members who were

linked to me maintained their family trees and incorporated information that was relevant to me and therefore affected my family tree, into them. I paid particular attention to idiosyncrasies that these members had included in their listings of dates, names and places.

This process continued until output data was of sufficient weight from which to draw conclusions. The additions of my ancestors were according to my knowledge and genealogical research preceding this study. When the present study showed my prior knowledge to be erroneous, I omitted the input data accordingly, or changed it if the output data suggested a more likely alternative. The logic here was to mimic the natural tree-construction process a user might employ when constructing the tree on the website for the first time, albeit I did it in a more structured and systematic manner following my chosen research method.

3.6 Ethical Considerations

Although numerous genealogical organisations have published ethics standards and codes of conduct pertaining to genealogical research (BCG, 2023b; Association of Professional Genealogists, 2023), these relate mostly to the conduct of professional genealogists and how it affects their clients. For a broader treatment of ethics in genealogy, I consulted Walters (2019).

Article 12 of the Universal Declaration of Human Rights (United Nations, 1948) recognises the right to privacy of all individuals, both in and of itself, and as a means of protecting themselves, their families and their respective reputations. Online genealogy is an activity with inherent privacy concerns. For this reason, many amateur genealogists do not construct their family trees online and do not participate in DNA testing. In order to allay fears of violation of privacy, websites such as Ancestry allow users to restrict what information of theirs (the user) is visible to other users. As emphasised by Walters (2019): "actually, *it is not compulsory to share your tree*". The implication here is that it is the prerogative of every user to decide for themselves what information they choose to share. Regardless of privacy settings, access to the personal details of living persons in a tree is restricted to the Ancestry.com member who constructed the tree.

I chose to use my own family history for this study, as any other data set could be intruding on others' privacy, the nature of online genealogy notwithstanding. The only individuals named were those of my deceased (presumed) ancestors. Even so, I referred to them in most instances by an alphanumeric code which I have described in Chapter 4. The paternity of a few of these ancestors is known to me due to DNA evidence only. In each case I maintained the anonymity of the inferred father to protect the privacy of his documented descendants.

As stated previously, it is the prerogative of every member to decide for themselves what information they choose to share on a site such as Ancestry.com. Likewise, they may also decide there is information they may not wish to know. In addition to what was presented in the previous paragraph, this also extends to sensitive information concerning adoptions, misattributed parentage and criminal ancestors, among many others (Walters, 2019: 74). Even if the sensitive information concerns someone who is no longer living, it may have a detrimental effect on their living relatives (Thomas, 2007: 228). Further to this point, Thomas (2007: 217), discussing privacy issues, emphasises that results of England censuses are only publicly released 100 years after they were conducted.

Ancestry.com allows members to upload their own photographs and documents and attach these to individuals in their family tree. However, it also allows other members to copy such member uploads to their own trees, which may lead to accusations of theft when copied without permission (Walters, 2019: 24). Although I have described such behaviour of Ancestry.com members in the chapters that follow, including the copying of information and other material from other trees, I maintained their anonymity so that they could not be accused of theft or plagiarism by third parties.

Further to the above, the present study was bound by the legal terms of use of Ancestry.com (Ancestry, 2024a). An analysis by Contreras et al. (2020) of these terms of use determined that the use of data from Ancestry.com (and similar sites) for academic research purposes is neither expressly allowed nor forbidden, even though much research has made use of the data. As per the terms of use, the intellectual property of Ancestry.com and its members was respected in this study, as was the privacy of Ancestry.com members and their immediate families.

Ethical clearance for this study was granted by the Ethics Review Committee of the Department of Knowledge and Information Stewardship at the University of Cape Town. The ethical clearance letter is included as Appendix C.

3.7 Data Analysis

I analysed my observations through IPA. IPA follows an iterative process of reading and re-reading transcripts – in this study, these were my observations as I had noted them down during the construction of my family tree – making additional notes, noting emergent themes, searching for connections across themes and looking for patterns across iterations in the process as advised by Smith, Flowers and Larkin, 2009, cited by Gorichanaz (2017: 4) and by Tuffour (2017: 4)).

I applied the iterative process in two ways. The first was to group output data pertaining to individual people in my tree as a whole or in specific family lineages. My observations and subsequent analysis thus pertained to how my tree-building (curation) activity affected those people or lineages. The second was to group output data by who performed the curation – in other words, who constructed family trees or uploaded documents which affected my own tree-building. In this way the hermeneutic circle is applied: by iteration, understanding of the parts (in the case of this study, individuals in any given tree) proceeds by understanding of the whole (in the case of this study, the tree as a whole) and understanding of the whole proceeds by understanding of the parts (Tuffour, 2017: 4). I also understand others' construction of their trees by understanding the construction of my own tree, and vice versa.

3.8 Research Rigour

Lincoln and Guba (1985) proposed four criteria by which to establish the trustworthiness of results of qualitative research: credibility (is it believable?), transferability (can it be applied in other contexts?), dependability (if repeated, will the same results be obtained?) and

confirmability (are the results independent of researcher bias?). Measures taken in this study to enhance research rigour according to these criteria are detailed below.

Rodriguez and Ryave (2002), in their description of SSO, note that studying experiences using interviews and surveys may result in data loss, as aspects of the experience that go unrecorded in the moment will not be remembered later. There is also the possibility of the researcher misinterpreting the participants and making incorrect inferences (Gorichanaz, 2017). These factors theoretically provide first person research methodologies with an advantage over third person methodologies. This approach allows for 'thick description' – detailed and contextualised descriptions of data (Babbie and Mouton, 2001), increasing transferability of results.

Several Ancestry.com features allow for a research audit trail (providing confirmability), such as a timeline of family tree edits, the ability to download GEDCOM files of the tree and the ability to write research notes on the profiles of persons included in the tree. However, taking advantage of these features may have risked the anonymity and privacy of Ancestry.com members. This study thus relied heavily on thick description.

The methodology I have chosen may uncover aspects of the genealogy research process that genealogists might not remember or admit to if they were interviewed or surveyed. However it cannot answer questions pertaining to genealogists' reasoning, justification or motivation behind their curatorial choices. Where I have speculated as to reasons, I have made that clear. For the most part, however, these questions are beyond the scope of the present study, but would constitute an organic follow-up to it to expand upon its conclusions.

3.9 Chapter Summary

In this chapter I have detailed the research methodology of this study and my reasons for my choices of approach and design. I have set out the ethical considerations involved and the measures I put in place to enhance research rigour.

Chapter 4: Presentation of Data

4.1 Introduction

During the period of April to September 2024, I made use of systematic self-observation to record my reactions and observations with regard to the research hints generated by the Ancestry.com website algorithm when I constructed a family tree of my ancestors on that platform. I also recorded my reactions and observations with respect to Ancestry.com member-constructed family trees on which those ancestors were included. I present my results narratively, with examples, according to themes that emerged from my data. To introduce my results, I first explain how I constructed my family tree for this study, demonstrating how I discovered the research hints and member-constructed trees mentioned above.

4.2 Constructing the Family Tree

I created a new family tree on Ancestry.com, adding myself and five generations of my direct ancestors. Each entry to the tree comprised their name and surname, and date and location data for their birth and death (where applicable). This information is drawn from personal research I have conducted previously as a person interested in their family history. In many instances, this information from my prior research was incomplete and/or approximate.

I have replicated the initial steps in my research process in figures 4.1 to 4.10. For privacy reasons, I have redacted information from some of the figures. I had a pre-existing Ancestry.com site login in order to construct family trees and a Worldwide membership subscription (Ancestry, 2025) on the website which grants me access to all records on Ancestry.com. Figure 4.1 shows the start of the tree-building process, at which point I clicked to add myself as the home person and filled in my details in the pop-up screen (figure 4.2), resulting in the nascent family tree (figure 4.3).

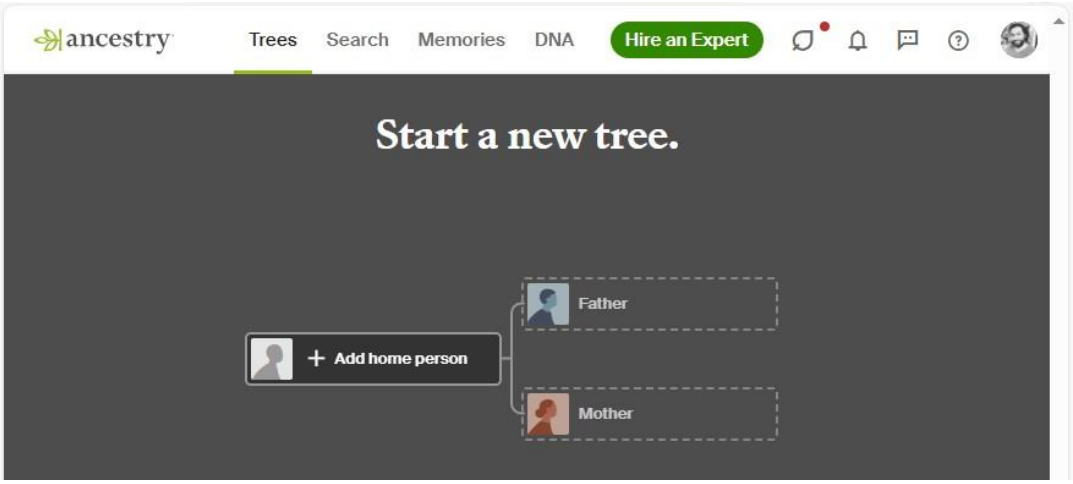

***Figure 4.1**. The start of the tree-building process on Ancestry.com.*

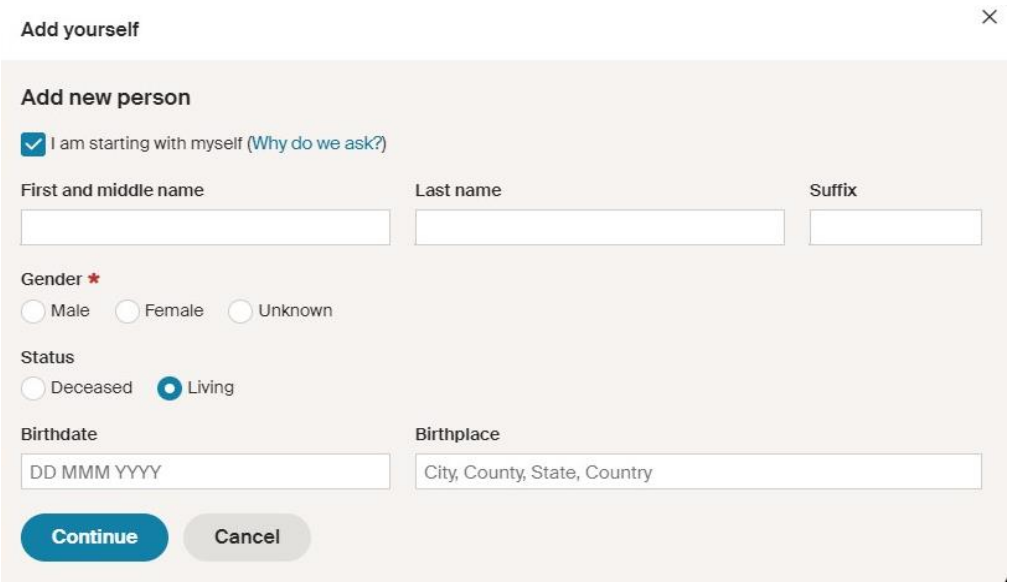

***Figure 4.2**. The pop-up screen that appears after clicking 'Add home person'..*

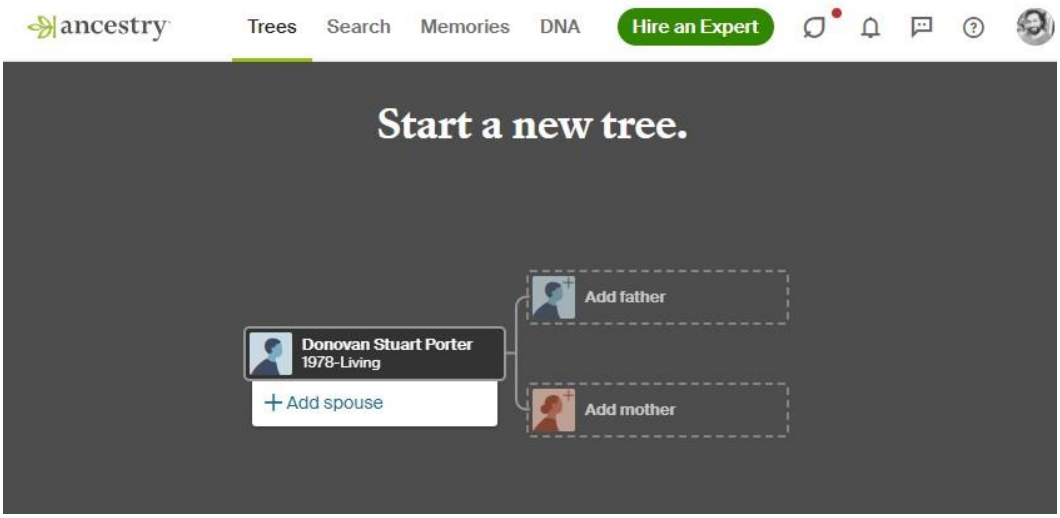

Figure 4.3*.The nascent family tree after adding the home person.*

I then added my father as the second person in the tree (Figure 4.4). There are two features of the tree to note evident in Figure 4.4. The first is the green leaf on my father's profile denoting 'hints' – available genealogical documents calculated to be relevant to my research by an Ancestry.com website algorithm. The second is a prediction of who my father's parents are. These predictions are based on existing trees (constructed by other Ancestry.com members) which include my father. Note also that at the top of the screen is a green link inviting me to "Upgrade to Pro Tools". Pro Tools is suite of features which requires an additional subscription (Ancestry, 2024d). Included is a "Tree Checker" which detects discrepancies in a member's family tree. Members cannot view which other members are subscribed to Pro Tools, therefore I could not make observations regarding it in this study. However, I make reference to it in the following chapter.

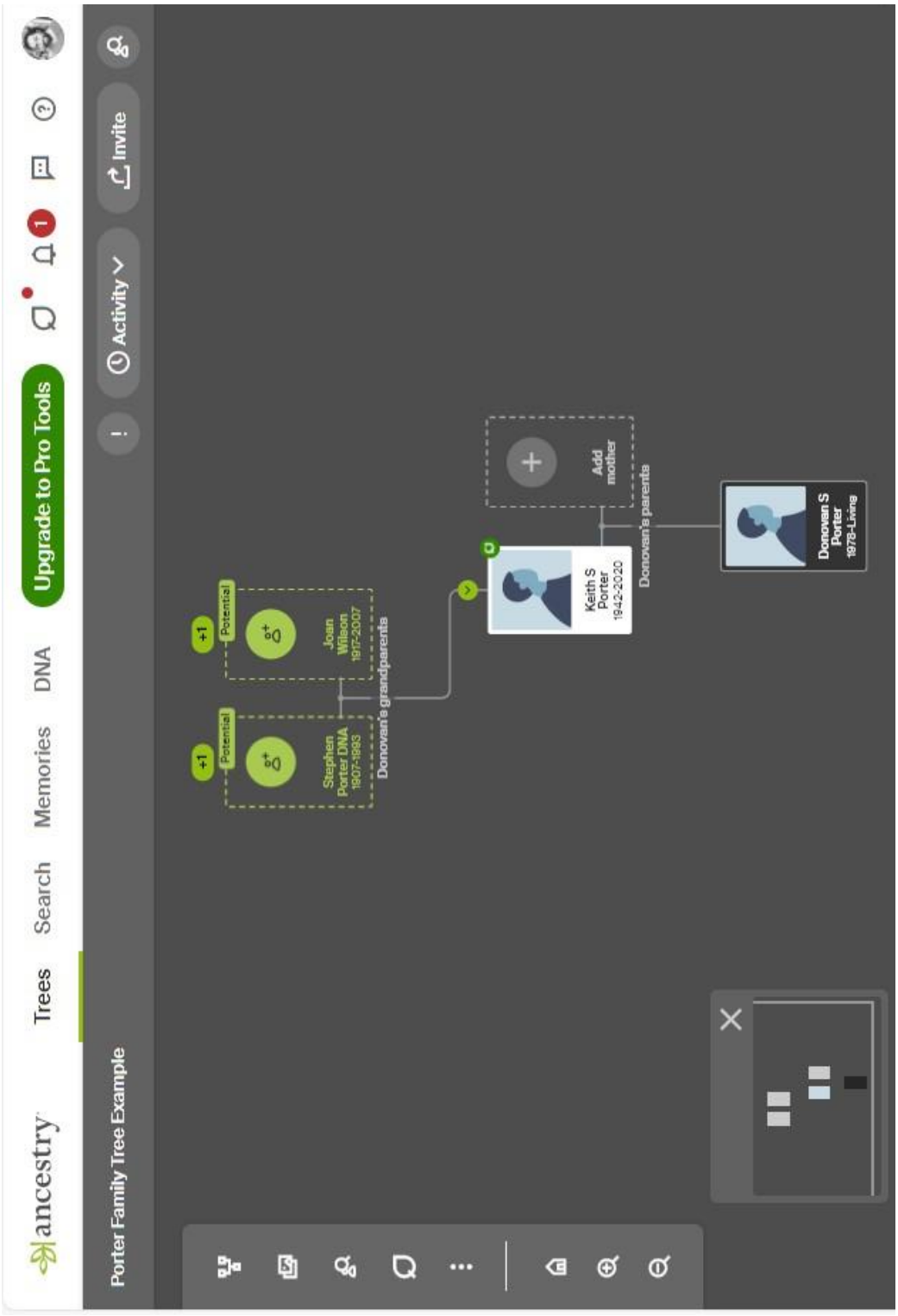

Figure 4.4. *The addition of my father to the tree.*

Note the leaf denoting hints, the Ancestry.com algorithm's prediction of the identity of my grandparents and the green link at the top of the screen inviting me to “Upgrade to Pro Tools”.

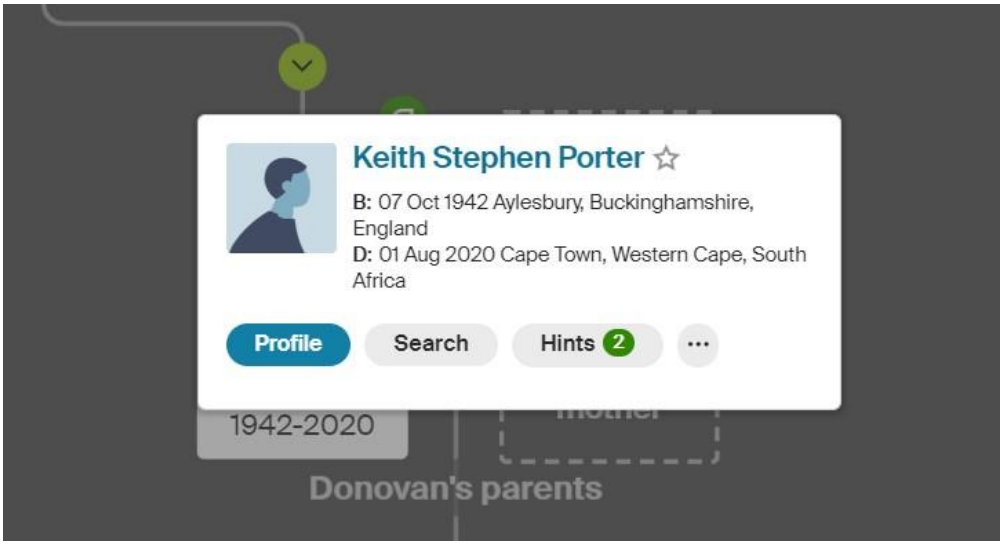

***Figure 4.5**. The pop-up that appears after clicking on my father in the tree.*

Clicking on my father's name produced a pop-up (Figure 4.5), from whence I could navigate to a tree profile page (created automatically when I added him to the tree), view the aforementioned hints (Figure 4.6) (which I can accept or ignore) or conduct a record search on Ancestry.com using his details that I have entered. The hints appear as a tab on the profile page. The 'Facts' tab includes the information I had entered about him and his close family – his parents, myself and my mother and 'LifeStory' presents this information in a website-generated narrative form. 'Gallery' shows any photographs or pictures I have uploaded or saved. Finally 'Explore' provides links to information and photographs Ancestry.com predicts I may be interested in. The search button on the pop-up also appears on the profile page and performs the same function, which is to do a search for records using the information I had entered previously.

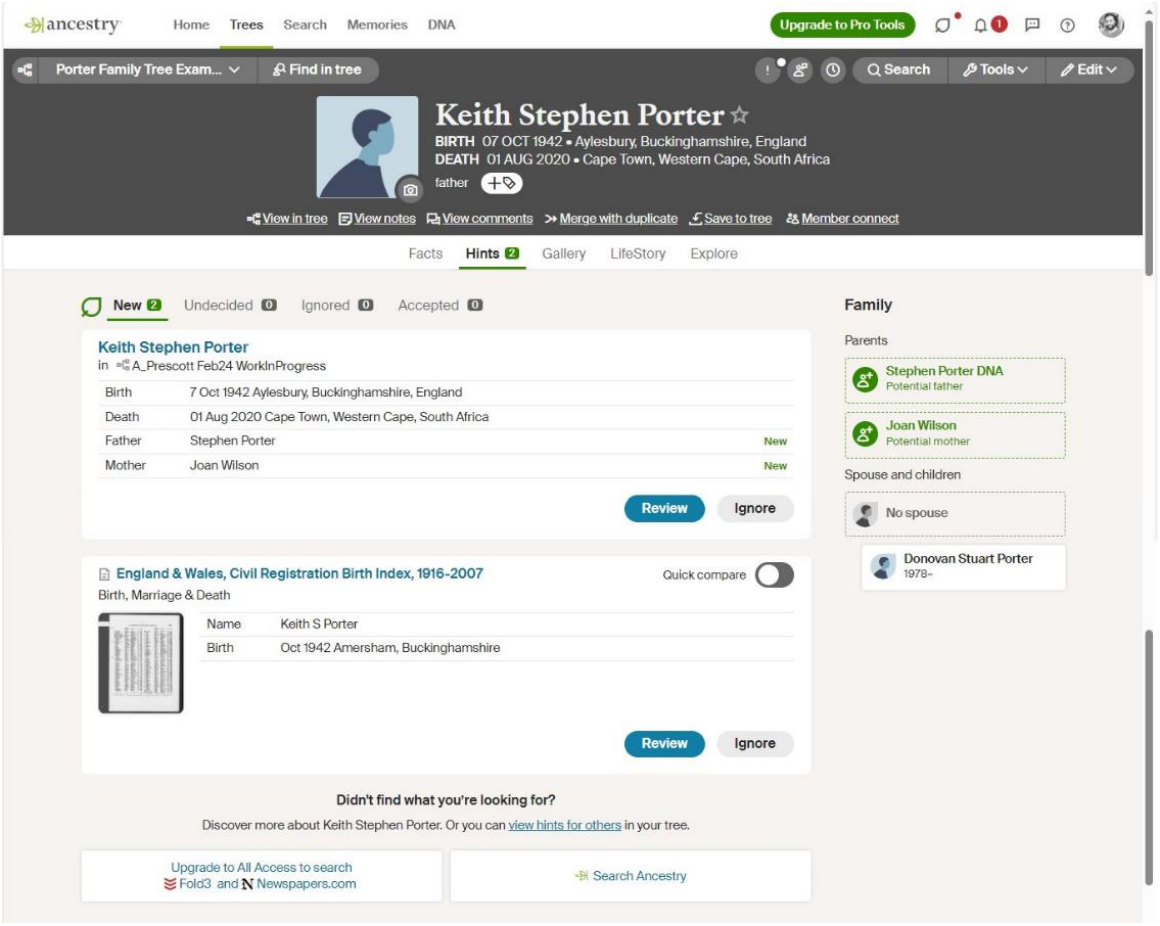

Figure 4.6. *My father's tree profile page, showing the hints tab.*

Note also the website-generated prediction of his parents on the right.

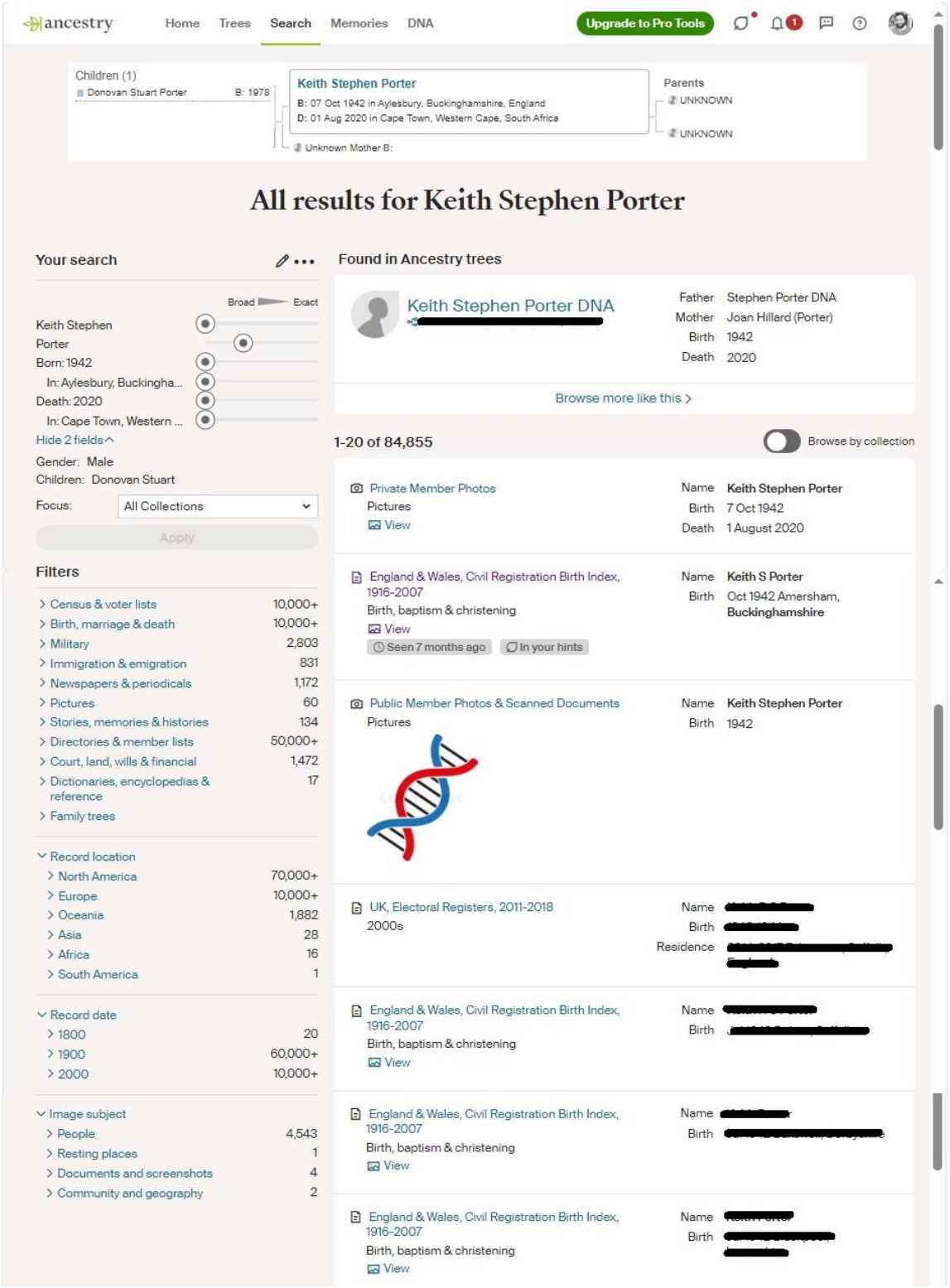

***Figure 4.7**. Record search results for my father.*

Note the filters available. I have redacted results not relevant to my father for privacy reasons.

In Figure 4.7 I have displayed the search results to emphasise the filters on the left side of the page. These relate to type of record, location, date and, for images, their subject. The search parameters (above the filters) are default settings, and with one exception, are set at their broadest specificity. I kept these settings to maintain the organic feel of the process. I also used the search function only to view a list of Ancestry.com member trees in which my ancestors appear. Thus, clicking on 'Browse more like this' (Figure 4.7: under the family tree search result) directs to a list of more family tree search results which feature my father (Figure 4.8). Again, the search parameters are set to a broad specificity. For each search I clicked on only one, rarely two, of these search results. For those searches where I clicked on two results, I recognised that the Ancestry.com algorithm had divided a list of family trees into two search results that were referring to the same person. Most often, the subject of the person I was performing the search on was the top search result. Clicking on 'View all' for the relevant result brings up the pop-up shown in Figure 4.9 which contains a list of family trees in which my father is included.

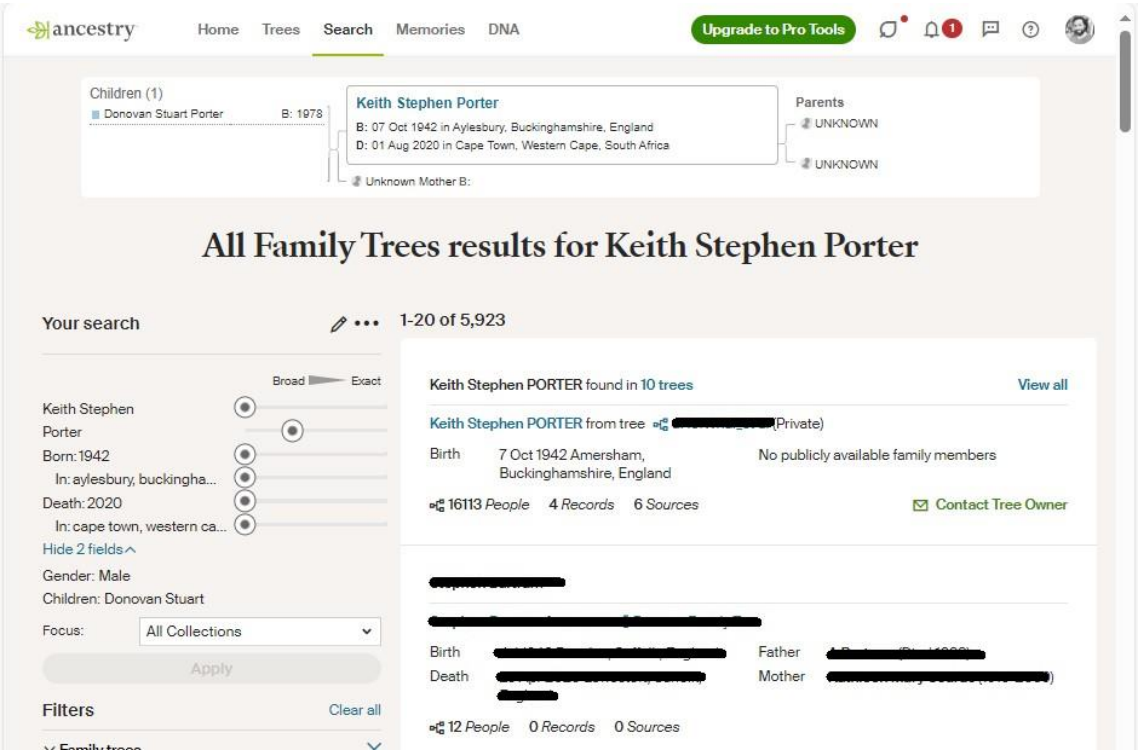

Figure 4.8. *Family tree search results for my father.*

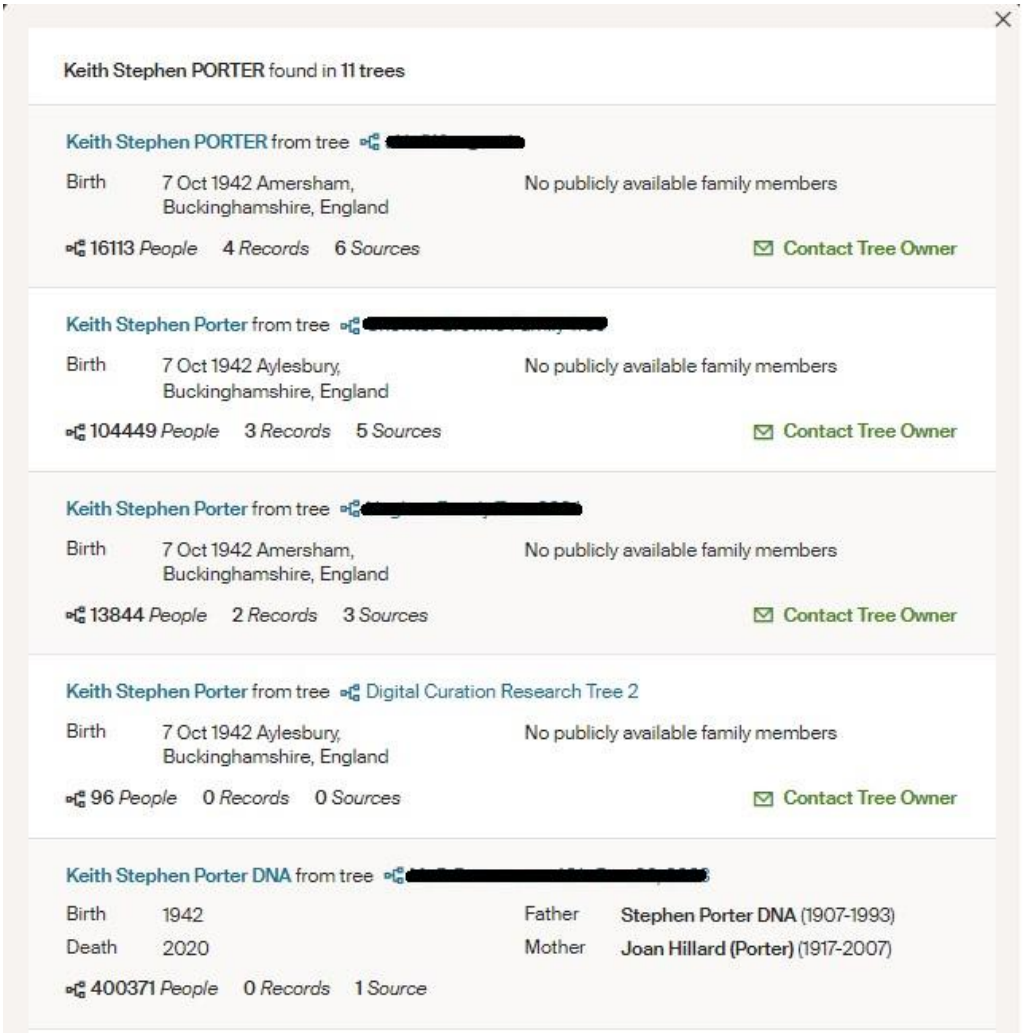

Figure 4.9. *The pop-up that appears after clicking on 'View all'.*

Note: I refer to this pop-up in the text as a tree summary page.

In this chapter, I shall refer to this pop-up (Figure 4.9) as a tree summary page. As can be seen in Figure 4.9, the first four trees listed in relation to my father are set to Private ('No publicly available family members') by the Ancestry.com member who constructed them – the fourth being the tree I constructed during the course of this study. The listing for each tree includes the number of people in the tree, the records attached to this specific person by the Ancestry.com member who constructed the tree and the number of sources they cited to obtain information on this person, such as records or other Ancestry.com member trees. For instance, the first tree listed contains 16113 people, has 4 records attached and cites 6

sources. The trees are listed in order of the number of records attached. For each tree in which one of my ancestors appeared, I recorded these pieces of information in a spreadsheet.

The tree summary page (Figure 4.9) was my point of access to view the trees to make observations for my research, such as use of MyTreeTags™ or detecting notable features of their trees. From each tree I could then view the member's account profile page, which contains information such as when they joined Ancestry.com and when they were last active on the site, which I also included in the spreadsheet.

4.3 Study Findings

In this study I refer to individuals in my constructed tree by an alphanumeric code consisting of a number according to the Ahnentafel system, followed by the initials of their name. I constructed an additional tree in Ancestry.com to demonstrate this system, which I have shown in Figure 4.10. The trees of other Ancestry.com members on which my ancestors appear, I have anonymised and named according to the context of which of my ancestors I first described them.

My observations regarding the research hints and member trees in which my ancestors appear constitute my research data, as they informed how I interpreted the curation of information by Ancestry.com and its members. I collected research hints from a total of 55 ancestors and 406 Ancestry.com member trees. My research data, available in appendices A and B), is summarised thematically in this chapter. Themes related to information I have entered are: place names (4.3.1), dates (4.3.2) and personal names (4.3.3). Themes related to the curation behaviour of Ancestry.com members are public vs private (4.3.5), citing sources (4.3.6), tree maintenance (4.3.7), comments and tags (4.3.8), copying user trees (4.3.9) and tree errors (4.3.10). Finally, themes related to the Ancestry.com website itself are inappropriate hints (4.3.4) and website oddities (4.3.11).

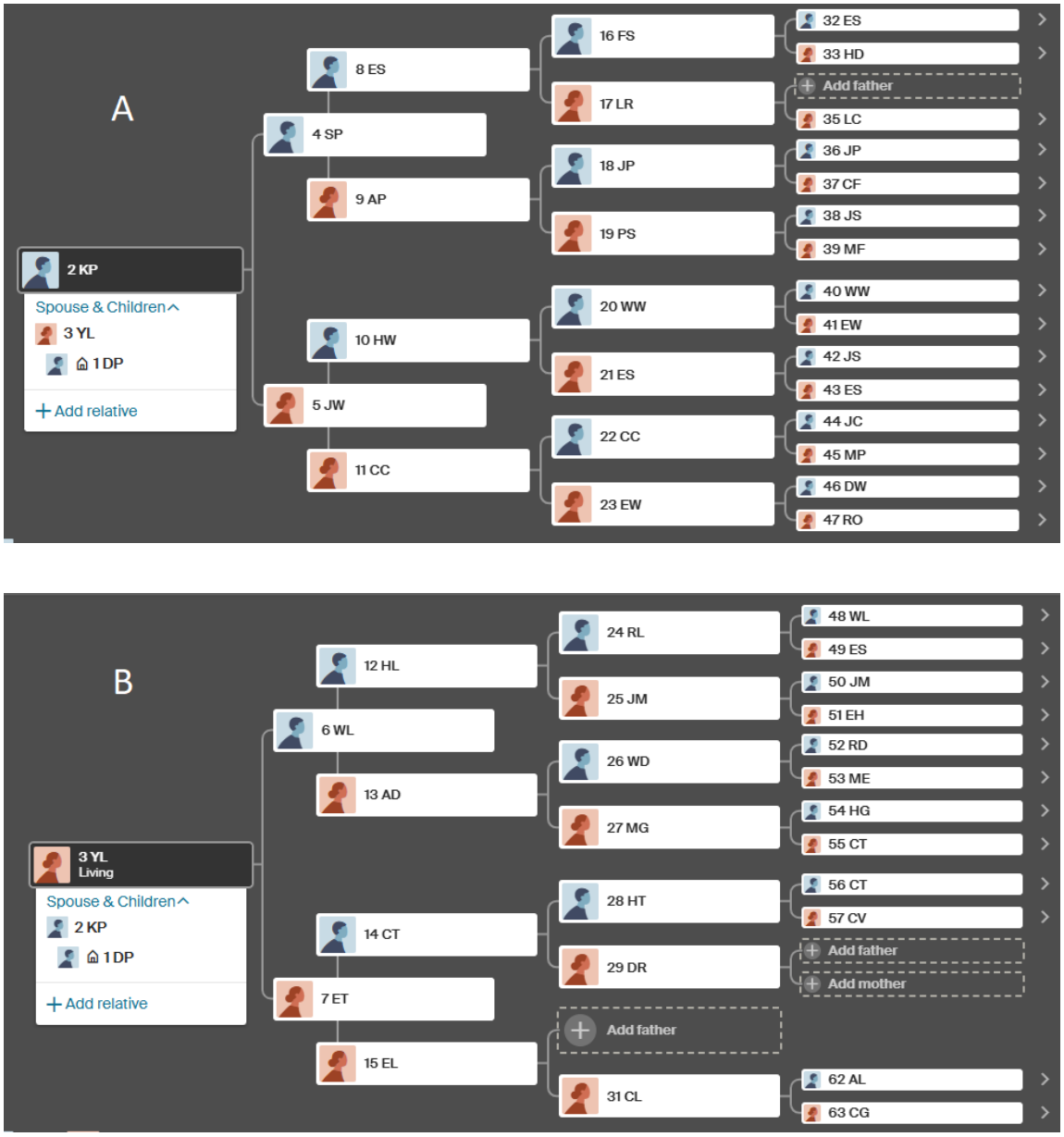

Figure 4.10. *The generations of my family tree that I have used in this study.*

Note: Here I demonstrate the alphanumeric code I use for naming, consisting of a number according to the Ahnentafel numbering system and each person's initials. As I can show only five generations simultaneously, I have omitted myself (1 DP) and separated the tree into the ancestors of A) my father, 2KP and B) my mother, 3YL.

4.3.1 Place Names

Place names in genealogy are highly informative, as names can change over time and administrative jurisdiction over an area can change. As a large proportion of my ancestry can

be traced to the south of England, many of the research hints on Ancestry.com taught me about the complexity of English administrative geography and how it has changed in recent history.

In 1928, my paternal grandfather, 4 SP, requested a certified copy of his record of birth. This certificate is now in my possession. It shows that he was registered in the district of Camberwell, sub-district Peckham South, in the county of London in 1907. His baptism, as per an Ancestry.com research hint, is recorded in the parish of Camberwell, while the Ancestry.com description of the document gives the full place of baptism as St Giles, Camberwell, Southwark, London. When entering information in a profile or when searching for records pertaining to a person, the Ancestry.com website attempts to simplify the search process or information input by providing a menu of location options (Figure 4.11), based on the searches and inputs of other Ancestry.com members. When entering Camberwell, the first option given is Camberwell, Surrey. Camberwell was indeed part of the county of Surrey until 1889, when it became part of the county of London, as it was when my grandfather was born in 1907. Early in my personal research, I once made this very error of selecting the option of Camberwell, Surrey – as I was unfamiliar with the history of boundary changes. Although I did not observe this exact error on member trees, I did see many instances of very precise places of birth and death which were anachronistic in their usage. One example of this was a member tree in which 5 JW appeared. The tree places a death in 1993 in 'Lewisham, Kent', even though Lewisham was moved from Kent to the new county of Greater London in 1965.

In the same vein, a hint – a South African Estate/Death Record issued by the Western Cape High Court – listed for my maternal grandfather (6 WL) attests to his death. Although the Western Cape did not yet exist as a province of South Africa at the time of his death in 1986, the use of the term was already being used in other contexts, such as administration of the courts. I observed that some member trees listing 6 WL gave his place of death as Western Cape, rather than Cape Province (which was the name of the province at the time of his death), an error also made Ancestry.com's transcription of the record.

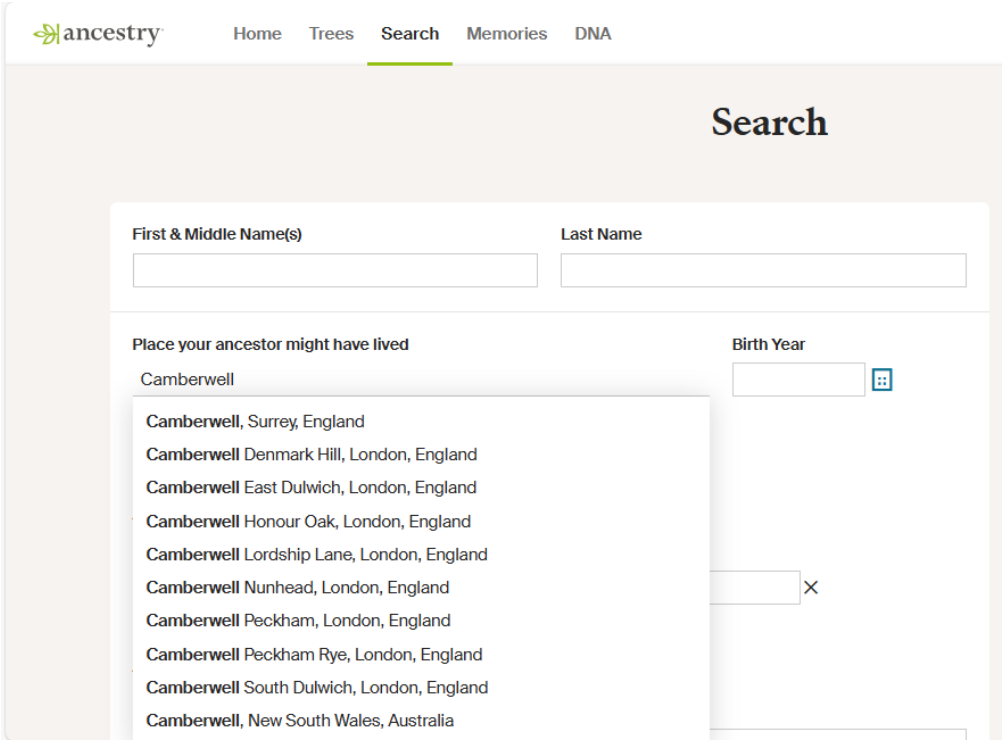

***Figure 4.11**. A drop down menu of location options when conducting a search for Camberwell.*

The events of 6 WL's death show another inconsistency in how Ancestry.com members have recorded it. 6 WL died in South Africa, where he was cremated. His ashes were flown to England by his daughter (my aunt) where he found his final resting place. However, Ancestry.com lists (as a hint) an additional death record for 6 WL that was created in England, giving the impression that England was where he had died. No image is attached to this record on Ancestry.com, so I was not able to view this record.

Returning to administrative geography, Ancestry.com hints for any given person differ in the exact information listed. For example, the England and Wales Civil Registration Birth Index records that the birth of 10 HW was registered in the district of Medway. In contrast, I found Brompton listed as 10 HW's town of birth on the 1891 English census. These two records do not contradict each other - Brompton was (and still is) a town in the district (now unitary authority) of Medway. Each record lists information according to the needs of the record type. Census returns rely on information given to the census worker by the person being enumerated and are thus susceptible to errors made by both parties. 24 RL, for example in

1861, 1871, 1891 and 1911 gave his place of birth as the village of Gretton, Gloucestershire, and in 1851, 1881 and 1901 he gave the neighbouring town of Winchcombe. It is incumbent on the genealogist constructing their family tree to be aware of local geography and consider all information available to them.

Where documents presented as Ancestry.com hints include the parents' address at the time of their child's birth, some users have incorporated that information into that person's place of birth in their trees. Alternatively, they may have based a person's place of birth on baptism records, so that users name the church where the child was baptised as part of their place of birth. I see a potential source of confusion here for someone trying to document their family history, as civil parishes (units of local government) were not identical to ecclesiastical parishes (units of Church administration, especially those of the Church of England). As per my comment above, genealogists need to be aware of these differences.

4.3.2 Dates

I have used census records to make an observation about 24 RL's place of birth, but they also offer insights into date of birth. When individuals were enumerated on English censuses, their age at their last birthday was recorded, but not their date of birth (The National Archives, n.d.), thus 24 RL's year of birth (1837) would be an estimate. Even so, the estimates given by the seven census records listed as hints for 24 RL's date of birth cover a range of years from 1835 to 1845. According to a baptism record on Ancestry.com, his baptism took place in 1838, so he was likely born in 1837. For comparison, the range of estimated year of birth for his wife, 25 JM, is narrower. She is recorded on six census records, with estimated year of birth from 1853 to 1857. Her baptism was in 1855.

I saw few Ancestry.com birth records that gave an exact date of birth for my ancestors I included on the tree I constructed for this study. Some Ancestry.com users have uploaded photographs of birth certificates and baptism records with such information – which the website algorithm offered to me as research hints. In the hints offered for 4 SP, I did find his exact date of birth (insofar as it was the date he celebrated as his birthday) on his Church of

England baptism record, his London workhouse records and the 1939 Electoral Register. Birth registration indices, which summarise registrations of births in a district per quarter, offered an approximate date of birth for some of my ancestors. I have cited the dates in such records in my own personal trees when records with more precise dates cannot be found. Finally, secondary sources offered as hints, such as user-produced entries on other websites, such as FindAGrave.com, sometimes included birth dates. I shall return to these in a later section.

4.3.3 Personal Names

I have observed the range of birth dates suggested by 25 JM's census and baptism records. The same records vary in the names under which she is recorded. Her baptism record refers to her as Julia Hannah and she is recorded as Julia H on the 1871 census. However, on the 1861, 1891, 1901 and 1911 census, she is simply Hannah. Her name on the 1881 census, perhaps due to a misunderstanding by the census worker, is Julianna. It was not uncommon for first names to be recorded in abbreviated or colloquial form, especially on census records. Other example are William who is recorded as W^{m} or Will^{m}, Richard as Rich^{d} and Henry as Harry. Specifically, 12 HL is recorded variously as Henry Richard (marriage record), Richard Harry (birth index) or Harry (1911 census and World War 1 service record). Julia Hannah and Henry Richard are most likely the 'correct' names of 25 JM and 12 HL respectively. I would only be sure of their correct names if I obtained copies of their birth certificates, which are not available on Ancestry.com.

I also observed variation in spelling of surnames on Ancestry.com. Some, such as Chandler/Chantler may be due to the misreading of handwritten documents, while others such as Stait/Staite and Morris/Morriss may be an error on the part of the official writing the document, such as on census records. However, variations such as Flexman/ Flaxman/ Flackston, Terhoven/ Terhoeven/ ter Hoeven and Darwin/ Darwinte/ Derwent imply some evolution of the name. I note here that on some older marriage records offered as hints, where the signature or mark of bride and groom are required, a crude X has been drawn, suggesting illiteracy. The marriage record of 26 WD and 27 MG is one such example.

I noted the use of Harry for Henry for two of my ancestors, whom I had both recorded on my tree as Henry. Four trees used both names for 10 HW in the form Henry (Harry) and another four (iterations of the same tree) used only Harry. Three member trees, by two users I inferred to be father and son based on their profiles, included 10 HW's year of birth as part of his name (between first and surname). On member trees, 12 HL was named variously as Henry (16 trees), Harry (8 trees), Richard Henry (9 trees), Henry (Harry) (3 trees) and Harry Richard (1 tree). Although some of this variation originates directly from source documents, some originates from information copied from other users' trees.

4.3.4 Inappropriate Hints

Of the research hints offered during my construction of my family tree, I considered only four to be entirely spurious. A newspaper entry announced the birth of a son in London to someone with the same name (but with an added middle initial) as 18 JP, who was my 2x great grandfather. The stated date and location of the birth were close enough to what I knew about where 18 JP and 19 PS were living at the time and when their youngest son (my grandfather's uncle) was born, that I could not dismiss the hint immediately. However, I concluded that although the details were similar, he was not the child whose birth was announced in the newspaper. I based this largely on where my grandfather's uncle was born and where his parents were living at that time.

A second spurious record was for a marriage of 20 WW, my 2x great grandfather in 1880. The confirmed marriage I know of his was 1877 in the same English county and his wife appears with him on the 1881 census record that I received as a hint. Unfortunately the Ancestry.com record is a transcript of the marriage record only, with no image to verify the transcribed information. If this is not a spurious record, it follows that this would be a bigamous marriage. While this is not impossible, I would need more information about the 1880 marriage to make that assessment.

The final two records were both listed as hints for 45 MP, my 3x great grandmother. One was a burial record for someone of the same name, albeit her pre-marriage surname. This person was buried in the English county of Norfolk, in the parish of Cockley Cley. However, according to a burial marker (member-submitted photo received as a hint) in the village of Charlwood, Surrey, 45 MP is buried here alongside her husband 44 JC. The other spurious record is that of a marriage between a woman sharing 45 MP's first name (but not her maiden name) and a man with a similar name to that of 44 JC.

I initially considered an 1851 census record for my 2x great grandfather 18 JP to be spurious. However, looking at the document itself, I could see that the Ancestry.com hint was correct, but his surname has been incorrectly transcribed from the document. It appears that despite the transcription error, the record is correctly indexed in Ancestry.com's database. In other cases I determined that I could not dismiss the hints entirely, in that there was a possibility that the hint was correct and my prior research which suggested otherwise was erroneous. The hints offered for 54 HG, for example, appear to relate to two different people who may have had similar but nevertheless different surnames. The spelling and transcription of both surnames were inconsistent in Ancestry.com's records and in the uploaded records of Ancestry.com members. When I looked at the list of hints from Ancestry.com that related to 54 HG, I started to doubt whether the information I had for 54 HG was correct – much of it originates from when I was less cautious in my research – and concluded that I may have conflated the details of two different people.

In fact, I found a hint for 13 AD from the Burial and Cremation Index on Ancestry.com that indicated that the date I had for her death was incorrect. In trying to make sense of the discrepancy, I determined that the source I had cited for the incorrect date was another Ancestry.com member tree whose information I had not evaluated for accuracy. Given that I now had a primary source for her date of death, I made the necessary corrections in my own research.

However, I found a large number of hints which were listed for people who were not the primary subject of the hint. For instance, a World War One service record for 6 WL was in fact that of his father - family members of service people are listed on the record. Similarly,

baptism records were listed as hints for the parents of the child in question. For example, hints for 36 JP contained three baptism records for three of his children with 37 CF, as well as a marriage record for another child. Despite the initial confusion that these records created, in most cases an image of the relevant document was available alongside the transcription to verify the attribution.

4.3.5 Public vs Private

Of the 406 Ancestry.com member trees I looked at, 314 were public and 92 set to private. Many members had created multiple trees, either as iterations of the same tree, or to emphasise different parts of their family. The ratio of public to private trees varies considerably across different family lineages and the number of trees on which any individual appears also varies considerably.

Ancestry.com member trees are set to public by default. Even so, only Ancestry.com members can view these trees and the details of living people are always private. Ancestry.com members can choose to make their trees private or even non-searchable – that is, searches will not display results from these trees. People's reasons for making their trees private are varied, but I can share my own experiences. In the course of my personal research, I have made two private trees on Ancestry.com, which I use to test hypotheses concerning the respective fathers of 4 SP and 15 EL. Their identities were unknown to me before I submitted samples for genetic ancestry testing. I am confident I now know who the former is and have identified someone I very strongly suspect to be the latter, though I still need further evidence to be sure. Although both these individuals are long deceased, they both have other living descendants. I would not make this information known in a space such as Ancestry.com, where it may be further disseminated by other members, without having first communicated with some of their respective descendants.

Further to my previous point, I have seen how my own personal information has been copied by other Ancestry.com members. I shall return to the copying of other member trees in section 4.12. Here though, I shall note that I felt uncomfortable seeing that other users who were only distantly or tangentially related to me had incorporated information such as the death of my father in 2020. Although my initial feeling that they had no right to do was an overreaction – I added the information to a public tree and the information was otherwise freely available – I understand why many members would keep private something which is so personal.

Finally, many of the private trees were among those with the most records attached to persons in the trees and cited the most sources, as evidenced by the numbers on the tree summary page (see also section 4.9). That is not to say that there is a definite relationship between the privacy settings of a tree and its reliability, although diligence in their research could motivate someone to keep their trees private because of the work that has gone into it. I also do not dismiss the possibility that some trees were created on another website or offsite in a specialised genealogy software program, and uploaded to Ancestry.com without citations.

4.3.6 Citing Sources

A citation on Ancestry.com is created when a member extracts information from a record in the database or a member tree hosted on the website. Members can also add their own citations of records not in the Ancestry.com database or information derived from other websites. When displaying tree matches for any person a member might search for (as in Figure 4.9), Ancestry.com lists the trees in order of the number of sources – such as records or other member trees – the member has made use of to assemble the information for the person being sought. I made the assumption that Ancestry.com intends the order of the list to be a useful and immediate method to gauge the relative reliability of any tree, even if the sources cited for the person in question are incorrect.

I was alarmed, though not surprised, that 93 of the 406 member trees I looked at did not cite any sources. Although the number of private trees is similar (92), there is no association or correlation between these numbers. Of the member trees that did cite sources, there were many whose only sources were other Ancestry.com member trees or information from other genealogical sites such as Geni.com, FindAGrave.com or Geneanet.org. The information on these sites that is offered as hints in Ancestry.com are equivalent to Ancestry.com member tree hints, in that they are at best secondary sources. This means therefore that they are constructing their trees with unverified information and not referring to primary sources. Ancestry.com draws hints from FindAGrave.com and Geneanet.org, but the Geni.com source is a member-added citation and other websites, not genealogy-related, can be cited in the same way.

4.3.7 Tree Maintenance

On a member's profile page, Ancestry.com displays when the member joined the site and (with very low precision) when last they were active on the site. Thus, I can see for example that in September 2000, my father joined Ancestry and at some stage started work on a tree (5 people) before perhaps leaving the family history research to me. The date he was last active on the site is simply stated as 'over a year ago'. It is possible that the entirety of his activity on Ancestry.com was on the day he joined. Thus 'over a year ago' could in fact be over 20 years previously. Two other family members, my father's cousin and my aunt (mother's sister), also began constructing trees, but did not make much progress before they halted. These are admittedly extreme examples of lack of maintenance of trees. Based on last active date, many other members whose trees included one or more of my ancestors, made a start on a tree, adding up to several hundred people before abandoning their trees. Of the 406 member trees I looked at, 82 were made by members who had been inactive on Ancestry.com for over a year.

It was gratifying to see other members who had joined the site 20 or more years ago and had been active on the website the day I had been studying their tree. Overall, 112 of the member trees I looked at were made by members who had been active on Ancestry.com that day. A further 13 were last on the site the day before and a further 49 in the previous two to six days. A limitation here is that unless I were to closely monitor changes made to member trees, I cannot see when other members updated their trees. I therefore used whether or not they were active on the website as a proxy activity.

Many Ancestry.com members maintain multiple trees. An extreme example is a user who maintains 302 trees on the site, of which only a few are private. As only one tree was pertinent to this study by containing one of my ancestors, I did not attempt to discover why this person maintained so many trees. In other cases the reason was much clearer. One member had two trees containing my ancestors and there was a third tree maintained by a relative, who I assumed to be his father, based on their names, ages and similar biographical information given. The title of each of these trees consisted of their family name, a date and a version number. In each case the version number corresponded to how many thousand people had been added to the tree. The most recent version of the tree contained in excess of 420k people. I looked at each of the three trees several times during my data collection and observed that the number of people on the tree had changed slightly each time. Given the size of the most recent version and the rate of growth from the previous versions, I concluded that the primary tree that these members use is likely offline and that they would upload another version at some point in the future.

Other examples of tree maintenance is versioning of tree names which consisted of a family name and a date. In several cases multiple dates had been added to the title, some of which also had the word ‘Copy’ in the title to differentiate them from similarly named trees. Another popular naming convention was simply variations of Tree 1, Tree 2, etc. or Tree (8)(1), Tree (8)(2), etc. Confusingly, some members appeared to have several versions or copies of a tree with the same name, with nothing to differentiate them. However, I determined that at least some of these listings on the tree summary page actually referred to duplications of the same person on one tree.

4.3.8 Comments and MyTreeTags™

Ancestry.com provides the functionality for members to write notes and comments on the profiles of people in the trees they have constructed (see an example in Figure 4.12), as a means to provide additional information about those people, or to express themselves regarding the research process. Notes are private, even on public trees, therefore I would not have seen them unless the member whose tree it was on, had specifically invited me to view them. Members can leave comments on other members' trees. Even so, I found use of comments, which are public on public trees, to be rare. Only six of the 314 public member trees that I looked at made use of the feature, and those that did, did so sparingly. I found the same to be true of MyTreeTags™ (Figure 4.13). While also a useful means of annotation, their use was limited to six member trees. I was not able to see which of the 92 private member trees made use of comments or MyTreeTags™.

Although Ancestry.com members did not make extensive use of these features, I observed another form of annotation. Many members uploaded icons as a profile picture for persons in their tree, signifying, for example, that the person was adopted, linked via a DNA match or served in the military. I also observed integration into the name of a person on their tree of birth dates and notations signifying how they are related to the person.

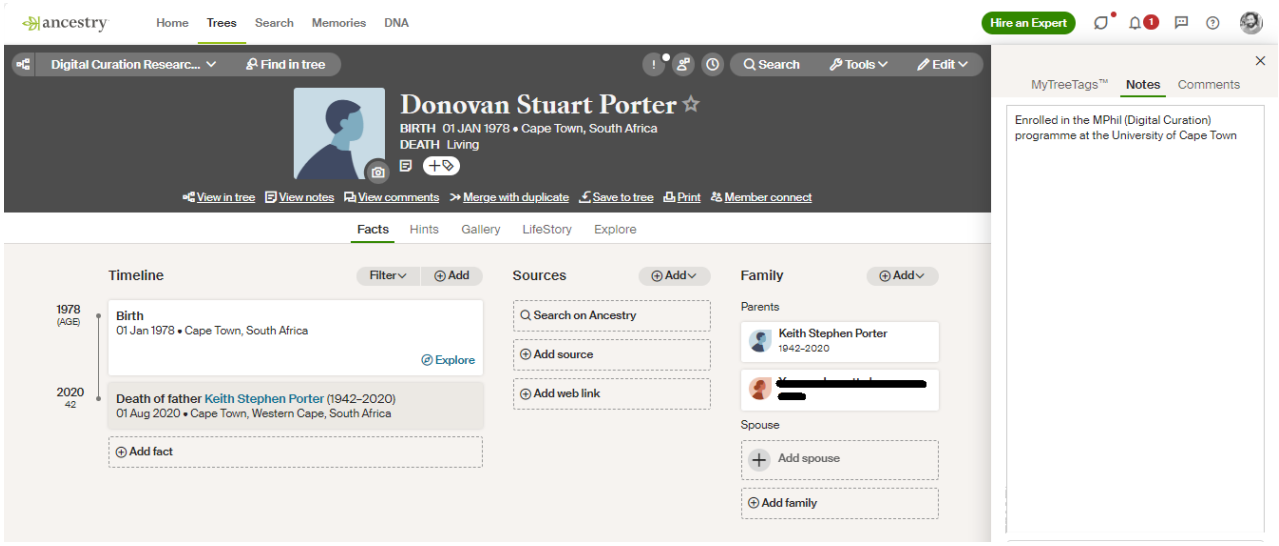

***Figure 4.12**. My profile on the tree I constructed for this study.*

I demonstrate the Notes function on the right-hand side of the screen.

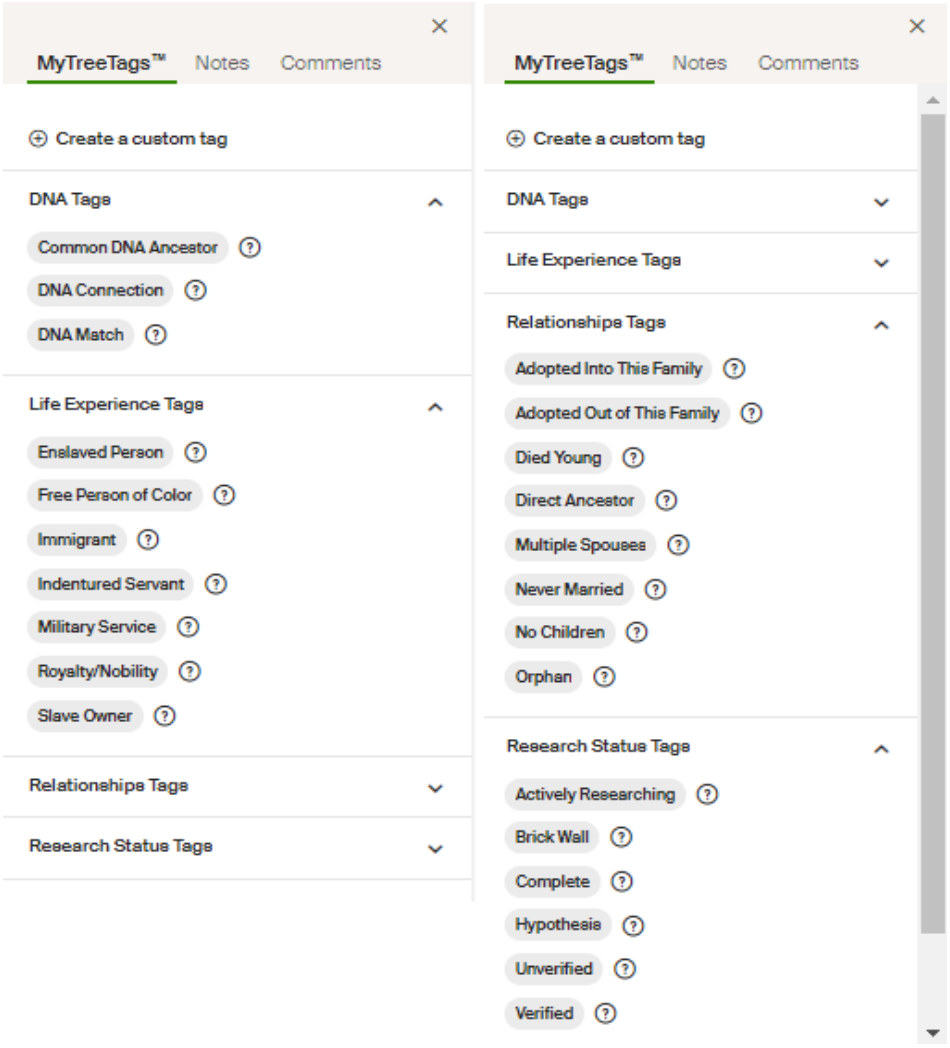

Figure 4.13. *A composite screen grab of MyTreeTags™ provided by Ancestry.com.*

4.3.9 Copying User Trees

During my earliest forays into researching my family tree, I was guilty of uncritically copying information from other members' trees. Thus, although I now take a critical view of these members for the manner in which they constructed their trees, I cannot claim any moral or intellectual high ground. I thought of this as I began seeing members citing my own tree as the source of their information.

Ancestry.com makes it very easy for members to copy extensively from other members' trees and their citation system shows from which trees members have copied. Indeed, 130 of the 314 public trees I viewed cited other Ancestry.com member trees or trees from other websites as a source for their information. It is possible that some Ancestry.com members have incorporated information from other trees into their own trees by other means, such as making notes with pen and paper, then manually adding the information. Again, I must add the caveat that I cannot see private trees (of which there were 92 in my research) and thus cannot make observations about the extent of tree copying in those trees.

One Ancestry.com member made use of quite particular notations to mark the position of many persons in their tree. For example, he annotated 13 AD as A/9/1/11/15/19(11 steps)%. When other members copied their tree, the member copying sometimes left the notation as is, not attempting to hide that they copied the information from elsewhere. I observed this phenomenon in trees containing 12 HL. On a subset of trees, his military rank was included as part of his name. It was clear that on at least some of these trees, this information had been copied from another tree. As I noted previously, members listing the first name of 10 HW and 12 HL as Harry or Henry (Harry) in their trees may also be indicative of how members have copied from others.

My own experience of copying from other trees occurred primarily when RootsWeb.com was more prominent and before I activated an Ancestry.com subscription. Much of my copying involved writing onto loose sheets of paper and rewriting into notebooks. Later I would rewrite everything again onto online trees, such as on Ancestry.com. At that stage I was only interested in names and dates and avoided nicknames and abbreviations or the annotations that other tree-makers used.

4.3.10 Tree Errors

It can be relatively easy to mistake two people with the same name who are living in the same town or village at the same time. However, I found some curious mistakes on the member trees I looked at. I made mention earlier of a member whose most recent tree had in excess

of 420k people. I discovered that he had conflated 48 WL, who was born in Gloucestershire, England, with someone of the same name who was born in New Brunswick, Canada, with the result that on his tree all ancestors of 48 WL were different to all other trees. A clue to this mistake may lie in the fact that he has copied much information from Geni.com. This genealogy website takes a different approach to Ancestry.com with respect to family trees. Instead of each member constructing their own trees, all Geni.com members add to and modify a single collaborative tree (Geni.com, 2024). I was a Geni.com member at the start of my genealogical journey, but ultimately left because of my frustrations with the collaborative tree, which made errors difficult to correct (fortunately Geni.com was not my primary genealogical website of choice). Thus, the error might not be his – rather, he did not verify the information he had copied.

In most cases, all the member-trees that members had copied from featured one or more of my ancestors – I had therefore listed those trees in the course of the study. In a few cases the tree that had been copied from was listed as having been subsequently deleted. Three trees (22 CC #25, 22 CC #72 and 22 CC#73) contained information about 22 CC that had been copied from trees that the Ancestry.com algorithm did not include in the list of trees that contained 22 CC. Further inspection of the three trees I listed confirmed 22 CC had been conflated with someone else of a similar name. A small cluster of trees, all constructed by the same Ancestry.com member had copied from the trees containing the incorrect information.

4.3.11 Website Oddities

I observed some oddities on Ancestry.com that I did not anticipate and had not noticed before in my use of the website. These observations pertain not to the activity of website members, but to how Ancestry.com displays information entered by members. A summary page of member trees (Figure 4.9) sometimes contained incongruous errors. Despite a clear line of documentation linking each person to a particular region to England, sometimes they were listed as living in other countries. The first instance I noted was a tree claiming that 13 AD was resident in Berkeley, California, United States in 1915. Intrigued by this assertion, I took a closer look at the information the user recorded for 13 AD. This user had not included the

incongruous information - their profile page for 13 AD correctly listed her residence in 1915 as Berdesley Green, Birmingham. Similarly, another user had the correct information for 13 AD. The summary page, however, records a 1915 residence for 13 AD in Mexico.

I observed this phenomenon on the summary pages of member trees when searching for 12 HL, 20 WW, 21 ES and 23 EW. After seeing this anomaly a number of times, I tried to determine a common denominator. In all cases, the location appearing incorrectly on the summary page corresponds to a location that has been entered incompletely on the profile page, specifically with country information omitted. The website, when creating the list of trees for the summary, appears to have misread the information on the profile page and created false information for the summary page.

4.4 Chapter Summary

While building my own family tree on Ancestry.com, I observed research hints offered by the website algorithm and member trees on the website in which my ancestors appeared. Research hints were generally relevant to the person in question, with very few hints that were incorrect. The Ancestry.com website itself sometimes displayed information from member trees that did not correspond with what the member had actually written. Most member trees were accessible to other members. Copying of information from other members was widespread, as was lack of citation of sources. Although many members had been active on the website shortly before I viewed their trees, many others had not been active for over a year.

Chapter 5: Discussion of Findings and Study Conclusion

5.1 Introduction

In this chapter I discuss my findings with reference to the literature, methodology and my data and according to my research questions which include study recommendations. I further discuss how online genealogy can be understood in terms of digital curation. I set out the limitations of the present study and make suggestions as to how future studies can build on it.

I have investigated the practice of genealogy in this study by viewing it through the lens of digital curation. I have analysed how Ancestry.com members (represented by myself) curate their genealogical information on the website, using the research hints provided by the website and the information from other member trees. Specifically, I sought to answer the following research questions:

RQ1: What is the ease with which a user can produce a family tree on Ancestry.com?

RQ2: To what extent are website-generated suggestions for the family tree correct or to what extent can they be verified?

RQ3: To what extent can users verify the information produced by other users for use in their own family trees?

RQ4: What suggestions can this study offer for the digital curation of these records which would result in a more accurate portrayal of a family tree?

I shall now discuss my findings in relation to each research question in turn.

5.2 Ease of Family Tree Construction

I address the question of the ease of constructing a family tree in Ancestry.com by referring to my observations concerning the maintenance of trees on the website, the number and size of trees constructed by Ancestry.com members and the access I had to Ancestry.com member trees. I relate these to the literature and my genealogy experience.

5.2.1 Maintenance of Family Trees

I observed that a large number (approximately 20%) of the member trees I looked at were constructed by members who had not been active on the website for over a year. Why would an Ancestry.com member cease activity on their trees and on the website? In Chapter 4 I gave the example of several family members, who made a tentative start on a family tree and subsequently did not continue. While my father and his cousin maybe started on theirs in a short-lived pique of wanting to solve a family mystery, my aunt had taken a DNA test with Ancestry.com, suggesting an alternative explanation for the lack of activity on the site: website members who take such tests are encouraged to construct their family trees as part of their genealogical experience. They might start a tree, but if that is not their primary interest, they might not continue for very long. But not all people who take DNA tests, whether with Ancestry.com or other providers, are interested in constructing their family tree. Discovering their 'ancestral regions', as Ancestry.com frames its ethnicity estimates (Ancestry, 2024b), may be their sole reason for taking a test, even if just as a novelty.

In this study I specifically investigated trees that shared members with my own tree. If I had encountered small family trees, which the Ancestry.com member had not continued to build upon, in the context of a DNA match, I would struggle to find a connection between my tree and theirs if we were not closely related. Their tree would therefore not have any value to me in my genealogical research. Erben (1991: 280) asserted that some genealogists are only interested in their family tree while the process is easy and the resources are accessible. Likewise, Darby and Clough (2013: 77) determined that genealogists begin with 'easy' research and proceed to more challenging tasks. For many genealogists, this progression from

easy to difficult culminates in a ‘brick wall’, where progress comes to a halt (Thomas et al, 2007: 29).

In my own family tree, the primary mysteries I have in the past sought to solve were the identities of the respective fathers of 4 SP and 15 EL. While DNA evidence eventually helped solve those mysteries (although I am not fully certain about 15 EL’s father), I first needed to positively identify their respective mothers (8 AP and 31 CL) on genealogical records before I could progress. In the case of 8 AP, it was, amongst other research, a careful process of tracing her across several English censuses; in the case of 31 CL, a serendipitous moment led me to investigate a specific record I may have otherwise overlooked. For that research, progress hinged on these breakthroughs. Other genealogists, perhaps, are not able to break through their brick walls.

In their model of the genealogy research process, Darby and Clough (2013) do not speak of an end to the process. Genealogy is seen as a continuous activity, as Yakel (2004) observed. Nevertheless, some genealogists may decide that their work has reached a natural conclusion, especially if their motivation was to solve a particular family mystery. While it is the most prominent, Ancestry.com is one among many websites offering a platform for online family tree construction and a database of genealogical records. It may be that some members have let their Ancestry.com trees stand idle whilst moving their primary site of family tree construction to another website or offline. In agreement with Roued, Castenbrandt and Revuelta-Eugercios (2023: 194), who found that genealogists select from a buffet of online resources, I once maintained family trees at Ancestry.com and MyHeritage.com, while researching at both sites, as well as FamilySearch.org. However, as I reached the final stage in Darby and Clough’s (2013: 81) model (“push back selected lines”) in my own personal research, I found it expedient to focus on Ancestry.com as my primary source of genealogical information.

It is worth noting in this regard that starting in September 2023, Ancestry.com began restructuring their membership subscriptions, resulting in some website features no longer being free to use (Estes, 2023). A further round of changes to subscription packages was introduced in February 2024. In June 2024 a new suite of features called Ancestry® Pro Tools

was introduced for which an additional subscription was required (Ancestry, 2024d). In addition, Ancestry.com have not introduced useful features, such as a chromosome browser, that experienced genealogists have expressed a wish for. Chromosome browsers, available through several other major DNA-testing companies, show where on each chromosome two DNA test-takers have inherited stretches of DNA from a common ancestor (Leonard, 2019: 72). The timing of the introduction of Pro Tools coincides with the period in my research in which I had begun to investigate member-constructed family trees; therefore it is unclear whether these changes would have had a direct effect on my observations regarding these trees. Nevertheless, it is possible that long-term user dissatisfaction with the website may be a factor in the lack of maintenance on many family trees.

The result of a lack of maintenance of certain trees is that for other members viewing them, those trees have become static, no longer being added to and no longer being corrected. As the Ancestry.com member is longer active on the site, there is a lower probability of feedback from that member if someone enquires about their tree.

5.2.2 Number and Size of Trees

Although most members maintained a single tree, a small group of members maintained a large number of trees, in one case 302. Based on the names given to them, the trees were either iterations of the same tree – with dates or other version number codes attached – or trees focussing on different parts of the member's ancestry. The former option constitutes a form of version control, although it would not be necessary to keep as many old versions as some members have kept. The format and descriptiveness of the names of the various tree versions were varied. While most seemed to follow a logical pattern, some appeared to confusing and chaotic, at least to me as an observer. In many cases it was not clear to me which version of the tree the member who had constructed it considered the most authoritative.

I observed a few very large trees, the largest of which contained in excess of 420k people. Who to include in a family tree is not always a simple question. Johnson Crow (2016)

advocates for erring on the side of inclusion, but also cautions that it is not necessary to actively research people who are not directly related to oneself. In my personal family trees I constructed to identify the respective fathers of 4 SP and 15 EL, I added as many people as possible so that I could throw my nets of speculation and theorising as wide as possible and in so doing remove the question of 'who not to include'. Likewise, Leonard (2019: 69) encourages genealogists to build their family trees "wide and deep" for best results.

With respect to the Ancestry.com member with the excessively large tree, each new DNA test taker to whom he is shown to be related would present him with a challenge to determine what the relationship is between them. In the light of my study, Durie's (2017: 3) and Erben's (1991: 279) cynicism regarding the motivations of genealogists with large numbers of people in their family trees rings true to an extent. My findings suggest to me that although there are stereotypes that do not accurately represent most genealogists, it may still be true that some individuals regard a large number of people in their family tree as an achievement. The difficulty for me when investigating this tree was determining how this member is related to me. In most cases, the individuals shared by both trees would be a clue, but in this case, the tree has incorporated so many individuals, including from both sides of my own family, that it is not useful to me for my tree construction.

5.2.3 Access to Trees

The ease with which I built my tree was dependant on the access I had to genealogical resources. I had a pre-existing Ancestry.com membership subscription which grants me access to all records on the website. There are lower subscription tiers which grant access to records from specific regions, depending on the regional Ancestry website used. For example, Ancestry.co.uk has a subscription option granting access only to British and Irish records (Ancestry, 2025). This aspect of access did not come into my study, but it is important to acknowledge nonetheless. Building my tree would also have been made easier by the tree verification functionalities in Pro Tools (Ancestry, 2024d), which I have discussed elsewhere.

Ancestry.com members blocking other members from viewing their trees by setting them to 'Private' has an impact on the ease with which someone can construct their tree. I found that the number and proportion of members setting their trees to 'Private' varied greatly between the various lineages making up my family tree. The impact of private trees on my availability to access information about any particular ancestor also varied greatly, the impact being greater if that ancestor appeared on very few trees and one or more of those were private. This raises the possibility that some of my ancestors left few descendants and would therefore have very few genealogists researching their lives. Although it would be mutually advantageous for these genealogists to collaborate, there is no guarantee that they would be willing or able to do so. I was fortunate when researching 15 EL and 31 CL a few years ago. I emailed a DNA match and she agreed to collaborate to determine how we are related. After months of sharing the discoveries we had each made, we concluded we are 4th cousins and each expanded our knowledge of our common ancestors and related family lineages.

Although private family trees result in the information contained therein not being immediately accessible, this is not necessarily an impediment to tree construction, as members may agree to collaboration as I described above, I return to the topic of private trees in section 5.4.2.

5.3 Accuracy and Veracity of Website-generated Suggestions

I address the accuracy and veracity of website-generated suggestions in Ancestry.com by referring to my observations concerning the inappropriate research hints provided by the website algorithm and also to the subjectivity and historicity of genealogical records. I related these to the literature and my genealogy experience.

5.3.1 Inappropriate Research Hints

As shown, I observed relatively few research hints suggested by the Ancestry.com algorithm that were entirely spurious. These were a newspaper report of a son born to 18 JP, a marriage of 20 WW and a marriage and a burial of 45 MP. In each of these cases, I had enough information about my ancestors to determine that the people to whom these records pertained were not respectively 18 JP, 20 WW and 45 MP. In addition to these, there were hints that were listed for people who were not the primary subject of the hint. For example, a World War One service record listed for 6 WL was actually for his father 12 HL. These could be easily recognised if the image of the record in question was available to verify the information, which was not always the case.

For a genealogist constructing their family tree for the first time, much more effort would be required to differentiate between two or more equally likely possibilities offered as research hints by the website. In the case of 54 HG, I concluded that the hints suggested by the Ancestry.com algorithm were referring to (at least) two different people whose life events I could not easily untangle. This particular situation is not a reflection of the quality of the algorithm: it is working with the information I had given it. In this study, the maximum information I gave the algorithm for each individual was their name and surname, and their date and place of their birth and death. Additional implied information, due to how I constructed my family tree for this study, was the information listed above for their spouse, parents and child. It is incumbent on the genealogist not to accept a hint from the algorithm unless it is supported by weight of evidence. Unfortunately, similar records of two people with the same name being born, getting married or dying in the same place at around the same time may be challenging for even the most experienced genealogists (Knight, 2024: 13). Lack of caution and further research in these instances may result in fictitious composite individuals being placed on family trees.

5.3.2 Subjectivity and Historicity of Research Hints

Although the records suggested by the Ancestry.com algorithm may be valid and correctly pertain to the person being researched, this is no guarantee that the information provided by those records is correct. I found many examples of personal names, place names or dates recorded differently across the various records pertaining to that person. The question arises of which of these records (if any) contains the correct information. Hatton (2021: 1) notes that experienced genealogists are aware that written records vary in reliability. The person writing the record may have made an error or the subject of the document may have provided the writer of the record with incorrect information (Durie, 2017: 4). As older records were written by hand, there is the question of whether the handwriting has been correctly interpreted and transcribed.

Ultimately, the narrative recorded by the genealogist is subjective, described by Durie (2017: 4) as a "process of diminishing deception". Although website algorithms may be efficient in generating suggestions, the genealogist needs to be cognisant of circumstances in which ambiguity may arise and recognise that their conclusions may be less certain as a result.

5.4 Veracity of Member-created Family Trees

I address the veracity of member-created family trees in Ancestry.com by referring to my observations concerning the attribution of sources by members, the degree of sharing tree data by members, the nature of the ingestion of new information by members and the extent to which members annotated their trees. I related these to the literature and my genealogy experience.

5.4.1 Attributions

Of the 406 member trees I looked at, 93 did not cite any sources. The online digital environment makes attribution a relatively simple process, because the attribution is created

automatically when a member extracts information from a record. Attribution is noted as a sign of good research practice and organisation in Thomas et al. (2007:160). In many academic studies of genealogy, attribution of sources is assumed. Yakel's (2004) investigation into genealogist's information-seeking behaviour asserts their role in information management, but makes no comment regarding diligence in attributing information. Likewise Darby and Clough (2013: 78), in describing the genealogical research process and access to records, assume that attribution of sources is occurring at each stage.

In contrast, Roued, Castenbrandt and Revuelta-Eugercios (2023: 200) discovered that only half of their participants saved links to their online sources, although this is not direct indication of degree of attribution. Willever-Farr and Forte (2014: 481) conducted interviews and studied message board posts on Ancestry.com and FindAGrave.com. They discovered that incorrect attribution and the resultant negative effects on genealogical research were a source of conflict between members on these sites. Lack of attribution contributes to poor research and it angers and frustrates fellow members on those websites.

While many trees in my study displayed no attributions at all, most had some, but were not comprehensive in their attributions, including those that copied extensively from other trees. A parsimonious explanation would be that the information in those family trees was compiled from a variety of sources, not exclusively Ancestry.com, because the lack of attribution may indicate that the information was typed in manually by the Ancestry.com member. As I discussed previously, this aligns with the buffet model of Roued, Castenbrandt and Revuelta-Eugercios (2023: 194) and my own experience in my personal research of using more than one website on which to search databases for records. Although members can write their own attributions on Ancestry.com to display that they found some information elsewhere they may ignore this option. When faced with the lack of attribution for a piece of information in a family tree, I consider that information unreliable.

5.4.2 Sharing of Tree Data

New Ancestry.com member trees are public by default and the onus is on the member to adjust privacy settings. Approximately 20% of the trees I listed had been made private by the member who had constructed them to make them inaccessible to other members. Naturally, I could not make any determinations on the contents of the trees, but I did note that some of the private trees were those that had the largest number of records attached to individuals in those trees. This is consistent with Walters' (2019: 20) observation that some genealogists, when they have worked hard on their trees, do not want to make that information freely available.

Genealogy, both in its subject matter and practice, is a deeply personal matter to the family historian researching their ancestors. It therefore carries inherent privacy concerns, especially when practised in an online space. Duff and Johnson (2003: 90) and Yakel (2004) both noted a high degree of collaboration and sharing among genealogists, although participants for both studies were sourced in part with the help of genealogical societies and therefore were more likely to seek and offer help to others. De Groot (2015: 114) also emphasises the predominant collaborative nature of genealogy and further notes that while Ancestry.com and other websites provide options for members to keep their information private, their business model is dependent on members sharing information amongst themselves on the website (De Groot, 2015: 114).

In chapter 4 I explained why I keep some of my own family trees private. My reasons include the copying of information from my tree (some pertaining to my deceased father) and respect for the acknowledged family of ancestors whom I have identified via genetic ancestry testing. Walters' (2019: 20) identifies several reasons why members may keep their trees private which closely mirror my own experience, which was: 1) information may be incorrect; 2) members put hard work into their research and might therefore not want to make it freely available; 3) members may have had a bad experience of other members taking information from a public tree without permission; and 4) they might be suspicious of who is assessing the tree. The final point, which indicates that people have broader privacy concerns, can also explain why genealogists may choose to keep their research offline and also why some people

do not wish to participate in ancestral genetic testing. With reference to points 2 and 3 above, Moore (2023: 9) points out that while professional genealogists are bound by a code of ethics and are obligated to comply with applicable laws on copyright and personal privacy, hobbyists may be oblivious to laws and ethics standards.

As I discussed in section 5.2.3, even if a tree is private, the member who constructed the tree may agree to one-to-one collaboration. According to my assessment of the characteristics of private trees, the member who requested information may benefit by way of verified and better sourced information

5.4.3 Ingesting Information

Of the 314 public trees I looked at, 130 cited other Ancestry.com member trees or trees from other websites as a source for their information. In some cases, the member who had copied information retained the annotations and idiosyncrasies of the member tree they had copied from, even though those features may not have relevance to the copier's tree, indicating that they had copied the information without verifying or contextualising it.

Willever-Farr and Forte (2014: 481) observed that many genealogists lamented Ancestry.com's approach to family tree construction. Ancestry.com markets the process as easy and highly accessible by promoting use of their research hints. However, the uncontrolled use of research hints by inexperienced genealogists resulted in participants in Willever-Farr and Forte's (2014: 481) study expressing their frustrations at what they saw as a lowering of standards on the website. Likewise, Knight (2024: 13) observed how online genealogy encourages the creation and dissemination of fake lives in family trees by "incautious researchers" who lump together several real individuals. Unfortunately, people who have made such mistakes can be reluctant to correct them, as observed by Moore (2023: 9). This study cannot speak to this last point, but it has found many tree makers not putting in the work of doing their own research, instead relying on the work of others. The implications for the veracity of member-generated information for other users is therefore that the reliability of many trees is questionable due to a reliance on unverified information.

5.4.4 Annotation of Trees

While the level of annotation of family trees is in itself not an indicator of the veracity of those trees, how annotations are used could be. For example, if a genealogist wrote a comment on their tree explaining how they came to a specific conclusion, I could infer a degree of reliability in that tree, in that it would show that the genealogist had considered and evaluated lines of evidence. However, the use of notes, comments and MyTreeTags™ on member family trees was minimal. I did however observe that members tagged persons on their trees in other ways. While I did not find that these alternative annotations affected my searches, I do not consider it aesthetically pleasing, mostly because information is included in an incorrect location and therefore poorly organised.

Apart from members using alternative annotation methods, such as using an icon in the place meant for a photograph, why else would they not make use of the tools provided by Ancestry.com? MyTreeTags™ is no longer a 'new' feature and members have had sufficient time to grow accustomed to it. Walters (2019: 74) questions whether adoptees should be singled out with a label to denote them in a family tree. By extension, the same argument could be put forward against the use of other MyTreeTags™ labels, such as Slave Owner or Enslaved Person.

The use of MyTreeTags™ and comments was so low, even among members who did make use of them, that I cannot make inferences about what this means for the veracity of the trees. Although I find the use of alternative annotations to be a sign of poor organisation, I cannot infer that the information within those trees is less reliable as a result.

5.5 Suggestions for Improved Curation for Ancestry.com

A discovery while carrying out this research was the Ancestry.com member tree 'summary page' (as I termed it) (Figure 4.9) displaying incorrect information when the member who had constructed the tree had omitted country-level geographical information. I only observed this for member trees and not for records hints, and I found no evidence for any member using

the incorrect summary information. However, it is possible that some did and the change resulted in the Ancestry.com algorithm not indicating that they are relevant to me. The fact that it occurred at all made me question whether similar issues in the workings of the Ancestry.com website and database were affecting the user experience in other ways that I did not observe. I demonstrated in Figure 4.2 and Figure 4.11 that place name data is entered in the format [Town, Province, Country] depending on the applicable geographical units. Thus a place is captured in a single data field, with the various geographical units separated by commas, as needed. It appears that without the country-level information, the website links the information entered to another listed location. Fixing the problem I highlighted require a change in the data fields Ancestry.com uses for data capture from members or a change in which it processes and displays that data.

Ancestry.com are transparent with respect to public member-constructed trees, as other members can see which records have been attached to individuals and from which trees the member copied. Even private trees have a degree of transparency, as the number of citations and records can be viewed for all searchable trees and it is possible to view when any given member joined the website. Ancestry.com also provides other functionalities in the form of notes, comments and MyTreeTags™. None of these pieces of information by themselves serve as an indicator of quality, but they allow other members to make an informed decision as to whether to copy information from those trees. Ancestry.com (2024c) do provide guidance on good research technique on their help pages. These are, however, not highly visible.

Potentially more useful than any of these is an option that requires an additional subscription: Ancestry® Pro Tools, which includes a feature that detects errors and discrepancies in family trees (Ancestry, 2024d). Quality assurance is thus treated as a value-added product and not a basic necessity. I am not a subscriber to Pro Tools – even if I were, I would not be able to view which members are subscribed to it without those members specifically granting me privileged access to their tree. Though understandable for privacy reasons, the existence of Pro Tools does present the opportunity for trees to be labelled with a quality grading or metric of some kind.

While in this section I have discussed improved curation for Ancestry.com in response to RQ3, in the following section I discuss the philosophical links between genealogy and digital curation that I observed during the course of my research.

5.6 Genealogy through the lens of Digital Curation

The objective of this study was to investigate online genealogy through the lens of digital curation. Both Dallas (2016: 423) and Sabharwal (2015: 11) discuss how 'digital curation' is understood differently across various fields of study, some of which have their own traditions of what constitutes 'curation'. Finding a common denominator between these possible meanings results in a need for a broader understanding of what constitutes digital curation. Yakel (2007: 338), discussing the various proposed definitions of digital curation, raises the possibility of it becoming "an umbrella term for digital preservation, data curation and digital asset and electronic records management."

For Sabharwal (2015:11), that umbrella covers a "digital curation workspace" which is broader than digital curation, narrowly defined. Dallas (2016: 440) notes: "Digital curation research… should focus on the study of actual practices in a diversity of contexts." The pragmatic digital curation that Dallas (2016: 441) describes occurs, among other places, on websites and cloud services, of which Ancestry.com and its services is an example on both accounts. Ancestry.com curates, in the custodial sense, a database of genealogically useful collections, while also storing the family trees constructed by its members. Ancestry.com members practise pragmatic digital curation in their construction of these family trees and collection of genealogical resources from the Ancestry.com website.

Friday (2014) and Roued, Castenbrandt and Revuelta-Eugercios (2023: 187) in developing their models of the online research process of genealogists did not explicitly link their work to digital curation, although information management is a strong theme in both studies and I see them as examples of the research into pragmatic digital curation to which Dallas (2016: 440) referred. Although the Curation Lifecycle Model (Higgins, 2008: 136) cannot be applied *in toto* to genealogy, models such as those of Friday (2014) and Roued, Castenbrandt and

Revuelta-Eugercios (2023: 187) can be developed further if an understanding of digital curation principles are applied to them.

The above notwithstanding, some of the actions of the Curation Lifecycle Model (Higgins, 2008: 136) can be linked to genealogy when viewing it through a digital curation lens. The large number of trees constructed by Ancestry.com members who had not been active on the site for over a year can be described as lack of *curation* and/or lack of *preservation planning*. Some members may have *migrated* their family trees to another website or offline – if so, they have not *disposed* of their Ancestry.com trees. In the case of members who have a large number of trees, some are instances of trees emphasising different parts of someone's ancestry – such trees may have been *transformed* from earlier trees.

New information, especially if it is suggested by the Ancestry.com algorithm, needs to be *appraised* before it is *selected* and *ingested*. Ancestry.com members need to *reappraise* the information on their trees if they are in doubt as to its veracity, *disposing* of it if necessary. They may want to use Pro Tools to do this, although it requires an additional subscription. In fact, subscription tiers on Ancestry.com constitute a form of *access* control. Another form of *access* control is the decision to make a tree private. However, should the tree be public, this allows for the *use* and *reuse* of the tree by other members, allowing them to *appraise*, *select* and ultimately *ingest* new information to their own trees.

Roued-Cunliffe (2017: 112) describes dates, places and names as metadata and in this study I have additionally interpreted notes, comments and MyTreeTags™ as metadata added by the Ancestry.com member. The problem I encountered of Ancestry.com incorrectly displaying place information on the summary page when the country had not been included, can then be interpreted as a problem of incomplete metadata. Likewise, where Ancestry.com members have not cited their information in family trees, I interpret as poor information management.

Historically, genealogy has had an ethos of access to and sharing of data so that it can be reused so that genealogists can help others research their family histories. The development of new technologies has broadened that access and increased the amount of digital genealogical material available. During the same period, new technologies have also

increased the amount of scholarly research data and digital curation evolved as a concept and activity to preserve and manage that data while also advocating for open access to and sharing of data to encourage reuse. Digital curation and genealogy share the same information philosophy and online genealogy incorporates digital curation, understood broadly. Viewing one in relation to the other in this study broadens the "digital curation workspace" of Sabharwal (2015:11) and provides genealogy with a platform from which it can be researched in academia.

5.7 Study Limitations

While auto-hermeneutics provided me with a degree of insight into Ancestry.com members' behaviour with regards to their construction of family trees, it also had limitations. Most importantly, beyond the information available publicly, my methodology could not provide me with insights into member trees that were kept private. I could not view MyTreeTags™ or comments on these trees and therefore I could not observe them in my study. I also could not view whether any information on these trees had been copied from other member trees or other websites. Taken together, it means I could not determine whether there were significant differences in curatorial activity or quality between private and public trees, which meant that my investigation is partially skewed towards public trees and the members who constructed them.

This study involved me making observations by looking at the details of the ancestors and relatives of Ancestry.com members. In that sense I was ethically limited to what I could report without the permission of Ancestry.com or its members, especially with regard to personal names. While anonymisation mitigated against this to a degree, I was aware of reporting too much information, lest a connection could be made to living persons. This also reduced the potential research rigour of the study.

5.8 Future Studies

The study limitations above present an opportunity for a future study that could approach the research problem through the use of different data collection methods, such as questionnaires, interviews or focus groups, as per Darby and Clough (2013: 75) or Roued, Castenbrandt and Revuelta-Eugercios (2023: 197). These methods can also be used to attempt to answer other questions that have presented themselves during the course of this study. For example, I used member activity on Ancestry.com as a proxy indicator of the extent of maintenance of member trees. The lack of member activity on a large number of trees led me to ask if some of these members were using a different genealogy website in preference to Ancestry.com or indeed constructing their family trees offline. It would be interesting to ask genealogists whether they use several websites simultaneously or cycle between them and discover what this means to their research.

Related to the above is the question of features available on Ancestry.com. Experienced genealogists such as Estes (2019) have lamented the lack of a chromosome browser. It would be useful to ask members if they wish for a feature such as this introduced, or indeed other features that are already available on other websites. In this study I noted that MyTreeTags™ and comments were used scarcely. Genealogists such as Estes (2019) have lauded this feature, so it would be interesting to ask Ancestry.com members why they do not use these features more widely, instead making use of unnecessary icons uploaded as pictures. Another useful feature introduced by Ancestry.com is Pro Tools (Ancestry, 2024d), which allows members to verify information on their family trees and eliminate errors and duplications. However, Ancestry.com requires an additional subscription fee to access this feature. It may be a fruitful avenue of research to determine if Pro Tools raises quality standards across the website or if the cost involved results in improvements being limited to trees constructed by members who already apply a high level of quality control.

On a more philosophical or even idiosyncratic note, I would be interested in extending this study to find out whether and how genealogists would be able to relate digital curation concepts to genealogy or vice versa, perhaps comparing the responses of genealogists already familiar with digital curation to the responses of genealogists who had not encountered it or

associated concepts. Such a study would assist in solidifying the bond I sought to make in the present study and its aim of furthering genealogy's standing in academia.

5.9 Study Conclusion

Genealogy is an activity that is mostly practised as a pastime, and is understudied from an academic research viewpoint (Hershkovitz, 2011:18). In this study I have investigated the practice of genealogy through the lens of digital curation. Through the investigation which involved answering my four research questions, I was able to conclude that the processes and activities involved in the construction of an online family tree can be interpreted and described using the concepts and activities constituting the digital curation of research data. While several models have been proposed to explain how genealogists conduct their research (Friday, 2014; Roued, Castenbrandt and Revuelta-Eugercios (2023: 187), a dialogue between these and curation lifecycle models may enable a description of online genealogy in digital curation terms.

References

23andMe Inc. 2025. *About us*. Available: https://23andme.com/en-gb/about [2025, January 17].

Ancestors Research South Africa. 2019. *Ancestors South Africa Family Tree & Genealogical Research*. Available: https://ancestors.co.za [2025, January, 17].

Ancestry. 2023. *Family Trees*. Available: https://support.ancestry.co.uk/s/family-trees [2023, January 25].

Ancestry. 2024a. *Ancestry Terms and Conditions.* Available: https://www.ancestry.com/c/legal/termsandconditions. [2025, January 23].

Ancestry. 2024b *Ancestry regions in AncestryDNA®*. Available: https://support.ancestry.com/s/article/AncestryDNA-Ethnicity. [2024, December 11].

Ancestry. 2024c *Ensuring your Tree is Accurate*. Available: https://support.ancestry.com/s/article/Ensuring-Your-Tree-is-Accurate. [2025, January 23].

Ancestry. 2024d *Ancestry® Pro Tools Membership*. Available: https://support.ancestry.com/s/article/Ancestry-Pro-Tools-Membership. [2025, January 24].

Ancestry. 2025 Start tracing your family tree. Available: https://ancestry.co.uk/offers/subscribe. [2025, February 7].

Association of Professional Genealogists. 2023. *Code of Ethics and Professional Practices*. Available: https://www.apgen.org/cpages/code-of-ethics. [2023, August 27].

Babbie, E. and Mouton, J. 2001. *The practice of social research*. South Africa: Oxford University Press.

Beagrie, N. 2008. Digital curation for science, digital libraries and individuals. The *International Journal of Digital Curation* 1(1):3-16. DOI: 10.2218/ijdc.v1i1.2.

Blom, K.A. 2022. Navigating narratives of genetic categorization at the frayed edges of identity. *New Genetics and Society* 41(4): 334-357. DOI: 10.1080/14636778.2022.2137121.

Board of Certification of Genealogists. 2023. a *Ethics and Standards*. Available: http://www.bcgcertification.org/ethics-standards. [2023, April 17].

Board of Certification of Genealogists. 2023. b *Genealogist's Code of Ethics*. Available: http://www.bcgcertification.org/ethics-standards/code/ [2023, August 27].

Booth, A., Sutton, A., Clowes, M. and Martyn-St James, M. 2022. Systematic approaches to a successful literature review. 3rd ed. London: Sage Publications Ltd.

Bottero, W. 2015. Practising family history: 'identity' as a category of social practice. *The British Journal of Sociology* 66(3):534-556. DOI: 10.1111/1468-4446.12133.

Brach, A.R. 2008. Exploring botanical history using ancestry.com. *Harvard Papers in Botany* 13(2): 245-251. Available: https://www.jstor.org/stable/41761908. [2023, May 11].

Chater, K. 2008. *How to trace your family tree in England, Ireland, Scotland and Wales.* London: Hermes House, Anness Publishing.

Chidester, D. 2000. *Christianity: a global history*. London: Allen Lane, Penguin Press.

Contreras, J.L., Schultz, K., Teerlink, C.-C., Maness, T., Meyer, L.J. and Cannon-Albright, L.A. 2020. Legal terms of use and public genealogy websites. *Journal of Law and the Biosciences* 7(1). DOI: 10.1093/jlb/lsaa063.

Creet, J. 2020a. Ancestry, inc. In *The genealogical sublime*. University of Massachusetts Press. 71-94. DOI: 10.2307/j.ctvxkn6xk.7.

Creet, J. 2020b. Confucius: The oldest genealogy in the world. In *The genealogical sublime*. University of Massachusetts Press. 26-37. DOI: 10.2307/j.ctvxkn6xk.5.

Creet, J. 2020c. The Mormons: doctrine and data. In *The genealogical sublime*. University of Massachusetts Press. 38-70. DOI: 10.2307/j.ctvxkn6xk.6.

Crow, R. 2002. *The case for institutional repositories: a SPARC position paper.* Available: http://sparcopen.org/wp-content/uploads/2016/01/instrepo.pdf. [2025, January 16].

Dallas, C. 2016. Digital curation beyond the "wild frontier": a pragmatic approach. *Archival Science* 16: 421-457. DOI: 10.1007/s10502-105-9252-6.

Darby, P. and Clough P. 2013. Investigating the information seeking behaviour of genealogists and family historians. *Journal of Information Science* 39(1):1-13. DOI: 10.1177/0165551512469765.

De Groot, J. 2015. On genealogy. *The Public Historian* 37(3):102-127. DOI: 10.1525/tph.2015.37.3.102.

De Groot, J. 2020. Ancestry.com and the evolving nature of historical information companies. *The Public Historian* 42(1):8-28. DOI: 10.1525/tph.2020.42.1.8.

Duff, W.M. and Johnson, C.A. 2003. Where is the list with all the names? Information-seeking behaviour of genealogists. *The American Archivist* 66(1): 79-95. Available: https://www.jstor.org/stable/40294218. [2023, April 17].

Durie, B. 2017. What is genealogy? Philosophy, education, motivations and future prospects. *Genealogy* 1(1):5. DOI: 10.3390/genealogy1010004.

Ellis, C., Adams, T.E. and Bochner, A.P. 2011. Autoethnography: an overview. *Historical Social Research* 36(4):273-290. Available: https://www.jstor.org/stable/23032294. [2023, May 22].

Erben, M. 1991. Genealogy and sociology: a preliminary set of statements and speculations. *Sociology* 25(2):275-292. Available: https://www.jstor.org/stable/42857622. [2024, December 19].

Estes, R. 2019. *Using Ancestry's new Match Grouping Dots aka MyMatchdots*. Available: https://dna-explained.com/2019/04/18/using-ancestrys-new-match-grouping-dots-aka-mygroupdots/ [2024, December 11].

Estes, R. 2023. *Ancestry to delete WorldConnect trees and shut down RootsWeb mailing lists –prepare now.* Available: https://dna-explained.com/2023/02/19/ancestry-to-delete-worldconnect-trees-and-shut-down-rootsweb-mailing-lists-prepare-now/ [2023, March 20].

FindMyPast, 2025*. FindMyPast*. Available: https://findmypast.com. [2025, January 17].

Fire, M., Chesney, T. and Elovici, Y. 2015. Data mining of online genealogy datasets for revealing lifespan patterns in human population. *ACM Transactions on Intelligent Systems and Technology* 6(2):1-22. DOI: 10.1145/2700464.

Friday, K. 2014. Learning from e-family history: a model of online family historian research behaviour. *Information Research* 19(4): 641. Available: https://files.eric.ed.gov/fulltext/EJ1050454.pdf. [2025, January 22].

Fukushima, K., Bourrier, K. and Parker, J. 2022. The lives of mistresses and maids: editing Victorian correspondence with genealogy, prosopography and the TEI. *Digital Humanities Quarterly* 16:1. Available: https://dhq-static.digitalhumanities.org/pdf/000595.pdf. [2023, May 11].

Gastmeier, T. 2018. A foreign familiarity. *Anthropologica* 60(2): 536-541. Available: https://www.muse.jhu.edu/article/713385. [2023, May 22]

GEDmatch. 2024*. GEDmatch*. Available: https://gedmatch.com. [2025, January 17].

Gene by Gene Ltd. 2025. *FamilyTreeDNA.* Available: https://familytreedna.com [2025, January 17].

Geni.com, 2024. *World Family Tree*. Available: https://www.geni.com/worldfamilytree/learn-more. [2024, November 29]

Goldman Sachs, 2025. *Ancestry CEO Deb Liu on leading the largest for-profit genealogy company*. Available: https://goldmansachs.com/insights/goldman-sachs-talks/deb-liu/ [2024, January 17].

Gomes, C, Palomo-Diez, S., Lopez-Parra, A.M. and Arroyo-Pardo, E. 2021. Genealogy: the tree where history meets genetics. *Genealogy* 5:98. DOI:10.3390/genealogy5040098.

Gorichanaz, T. 2017. Auto-hermeneutics: A phenomenological approach to information experience. *Library & Information Science Research* 39:1-7. DOI: 10.1016/j.lisr.2017.01.001.

Graham, P.S. 1995. Requirements for the digital research library. *College & Research Libraries* 56(4):331-339. DOI: 10.5860/crl_56_04_331.

Gramps, 2022. *Research, organize and share your family tree with Gramps*. Available: https://gramps-project.org/blog/ [2023, April 17].

Guerrini, C.J., Robinson, J.O., Bloss, C.C., Brooks, W.B., Fullerton, S.M., Kirkpatrick, B., Lee, S.S.-J. and McGuire, A.L. 2022. Family secrets: experiences and outcomes of participating in direct-to-consumer genetic relative-finder services. *The American Journal of Human Genetics* 109:486-497. DOI: 10.1016/j.ajhg.2022.01.013.

Halley, J. 2011. Genealogy: a tale of two families and a cat. *Cultural Studies – Critical Methodologies* 11(3):285-289. DOI: 10.1177/1532708611409545.

Hatton, S.B. 2021. Genealogy's assumptions about written records and originality. *Genealogy* 5:21. DOI: 10.3390/genealogy5010021.

Haynes, D. 2018. *Metadata for Information Management and Retrieval* 2nd Ed. London: Facet. DOI: 10.29085/9781783302161.

Hershkovitz, A. 2011. Leveraging genealogy as an academic discipline. *Avotaynu* 26(3):18-23. Available: http://genealogy.arnononthe.net/files/genealogyAcademization.pdf. [2023, April 9].

Hershkovitz, A. 2012. A suggested taxonomy of genealogy as a multidisciplinary academic research field. *Journal of Multidisciplinary Research* 4(3):5-21.

Higgins, S. 2008. The DCC Curation Lifecycle Model. *The International Journal of Digital Curation* 3(1): 134-140. DOI: 10.2218/ijdc.v3i1.48.

Higgins, S. 2012. The lifecycle of data management. In: *Managing Research Data*. G. Pryor, Ed. London: Facet. 17-45. DOI: 10.29085/9781856048910.003.

Hjorthén, A. 2022. Reframing the history of American genealogy: on the paradigm of democratization and the capitalization of longing. *Genealogy* 6:21. DOI: 10.3390/genealogy6010021.

Hoeve, C.D. 2018. Finding a place for genealogy and family history in the digital humanities. *Digital Library Perspectives* 34(3):215-226. DOI: 10.1108/DLP-11-2017-0044.

Intellectual Reserve, 2021. *About GEDCOM*. Available: http:llgedcom.io/about/ [2023, April 17].

Intellectual Reserve, 2024a. *FamilySearch Library*. Available: https://www.churchofjesuschrist.org/learn/familysearch-library/ [2025, January 17].

Intellectual Reserve, 2024b. *FamilySearch 2024 Genealogy Highlights*. Available: https://www.familysearch.org/en/newsroom/familysearch-2024-genealogy-highlights/ [2025, January 17].

Intellectual Reserve. 2024c. *Welcome to RootsTech.* Available: https://www.familysearch.org/en/rootstech/ [2024, December 17].

Johnson Crow, A. 2016. *Who should you include in your family tree?* Available: https://www.amyjohnsoncrow.com/include-family-tree/ [2024, December 11].

Kaplanis, J., Gordon, A., Shar, T., Weissbrod, O., Geiger, D., Wahl, M., Gershovitz, M., Markus, B., Sheikh, M., Gymrek, M., Bhatia, G., MacArthur, D.G., Price, D.L. and Erlich, Y. 2018. Quantitative analysis of population-scale family trees with millions of relatives. *Science* 360:171-175. DOI: 10.1126/science.aam9309.

Kaschula, R.H. 1999. Imbongi and griot: toward a comparative analysis of oral poetics in Southern and West Africa. *Journal of African Cultural Studies* 12(1): 55-76. DOI: 10.1080/13696819908717840.

Knight, T.D. 2024. Genealogical ethics in the United States and the popularization of genealogical research in the digital age. *Genealogy* 8:78. DOI: 10.3390/genealogy8030078.

Kriesberg, A. 2017. The future of access to public records? Public-private partnerships in US state and territorial archives*. Archival Science* 17:5-25. DOI: 10.1007/s10502-016-9268-6.

Lee, S. and Park, J.H. 2019. Quality over quantity: a lineage-survival strategy of elite families in premodern Korea. *Social Science History* 43: 31-61. DOI: 10.1017/ssh.2018.38.

Leonard, M. 2019a. Understanding the principles of DNA testing for genealogy. In *Tracing your ancestors using DNA.* G.S. Holton, Ed. Barnsley, England: Pen &Sword. 32-59.

Leonard, M. 2019b. atDNA tests. In *Tracing your ancestors using DNA*. G.S. Holton, Ed. Barnsley, England: Pen &Sword. 60-103.

Leonard, M., Macdonald, A.F. and Holton, G.S. 2019. Why use DNA testing for genealogy? In *Tracing your ancestors using DNA*. G.S. Holton, Ed. Barnsley, England: Pen & Sword. 1-12.

Lincoln, Y. and Guba, E. 1985. *Naturalistic inquiry*. Beverly Hills, California: Sage Publications.

Living DNA Ltd. n.d*. Living DNA*. Available: http://livingdna.com [2025, January 17].

Lynch, C.A. 2003. Institutional repositories: essential infrastructure for scholarship in the digital age. *Portal (Baltimore, Md)* 3(2): 327-336. DOI: 10.1353/pla.2003.0039.

Macdonald, A.F. and Cleary, J. 2019. Y-DNA tests. In: *Tracing your ancestors using DNA.* G.S. Holton, Ed. Barnsley, England: Pen &Sword. 104-146

Mahuika, N. 2019. A brief history of Whakapapa: Māori approaches to genealogy. *Genealogy* 3: 32. DOI: 10.3390/genealogy3020032.

Mayernik, M.S. 2020. Metadata. *Knowledge Organization* 47(8): 696-713. DOI: 10.5771/0943-7444-2020-8-696.

Meyer, L. 2020. Family history: fact versus fiction. *Genealogy* 4:44. DOI: 10.3390/genealogy4020044.

Mitchell, B.A and Kim, B. 2024. Amateur family genealogists researching their family history: a scoping review of motivations and psychosocial impacts. *Genealogy* 8: 3. DOI: 10.3390/genealogy8010003.

Moore, S.M. 2023. Ethical dilemmas and family history: a psychological approach. *Genealogy* 7: 67. DOI: 10.3390/genealogy/7030067.

MyHeritage Ltd, 2025. *MyHeritage.* Available: https://myheritage.com. [2025, January 17].

Neilson, L.C. 2022. Accessing primary sources for marketing and advertising history research from family history websites. *Journal of Historical Research in Advertising* 14(2): 281-291. DOI: 10.1108/JHRM-10-2021-0055.

O'Brien, M. 2020. *Is RootsWeb still active? Will Ancestry shut it down?* Available: https://datamingdna.com/is-rootsweb-still-active-will-ancestry-shut-it-down/ [2025, January 17].

Oliver, S.T. 2021. My grandfather's face: a black contemplative educator's discovery of unknown relatives through DNA. *Genealogy* 5:103. DOI: 10.3390/genealogy5040103.

Otterstrom, S.M., Bunker, B.E. and Farnsworth, M.A. 2021. Development of the genealogical FamilySearch database and expanding its use to map and measure multiple generations of American migration. *Genealogy* 5:16. DOI: 10.3390/genealogy5010016.

Poole, A.H. 2016. The conceptual landscape of digital curation. Journal of Documentation 72(5):961-986. DOI: 10.1108/JD-10-2015-0123.

Putman, D.L. and Cole, K.L. 2020. All hail DNA: the constitutive rhetoric of AncestryDNA advertising. *Critical Studies in Media Communication* 37(3):207-220. DOI: 10.1080/15295036.2020.1767796.

Raeburn, T. Liston, C., Hickmott, J. and Cleary, M. 2017. Liverpool "lunatic asylum": a forgotten chapter in the history of Australian health care. *Collegian* 25:347-353. DOI: 10.1016/j.colegn.2017.08.003.

Rodriguez, N and Ryave, A. 2002. *Systematic self-observation*. Thousand Oaks, California: Sage Publications. DOI: 10.4135/9781412986076.

Roued, H., Castenbrandt, H. and Reveulta-Eugercios, B.A. 2023. Search, save and share: family historians' engagement practices with digital platforms. *Archival Science* 23:187-206. DOI: 10.1007/s10502-022-09404-4.

Roued, H. and Klareld, A. 2025. Multidisciplinary research on family historians. In *The Routledge companion to libraries, archives, and the digital humanities.* I.G. Russell and G. Layne-Worthey Eds. London, New York: Routledge. 247-259. DOI: 10.4324/9781003327738-21.

Roued-Cunliffe, H. 2017. Collection building amongst heritage amateurs. *Collection and Curation* 36(3): 108-114. DOI: 10.1108/CB-01-2017-0003.

Sabharwal, A. 2015. Digital curation lifecycle in practice. In: *Digital curation in the digital humanities: preserving and promoting archival and special collections*. United Kingdom: Elsevier Science & Technology 95-108. DOI: 10.1016/B978-0-08-100143-1.100005-2.

Saunders, R.A. 2024. Genealogical journeys, geographical imagination and (popular) geopolitics in *Who Do You Think You Are*? *Social & Cultural Geography* 25(5): 698-717. DOI: 10.1080/14649365.2023.2177714.

Schurink, W.J., Schurink, E.M. and Fouché, C.B. 2021. Thematic inquiry in qualitative research. In: *Research at grass roots*. C.B. Fouché, H. Strydom and W.J.H. Roestenburg, Eds. 5th ed. Pretoria: Van Schaik. 289-310.

Shiue, C.H. 2016. A culture of kinship: Chinese genealogies as a source for research in demographic economics. *Journal of Demographic Economics* 82:459-482. DOI: 10.1017/dem.2016.24.

Sleeter, C. 2020. Critical family history: an introduction. *Genealogy* 4:64. DOI: 10.3390/genealogy4020064.

Smith, J. A., Flowers, P., and Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method, research*. London, England: Sage Publications

Stallard, M. and De Groot, J. 2020. "Things are coming out that are questionable, we never knew about": DNA and the new family history. *Journal of Family History* 45(3):274-294. DOI: 10.1177/0363199020906853.

Starasta, M. 2009. Rootsweb: finding our roots together. *Reference Reviews* 23(4):54. DOI: 10.1108/09504120910958692

Sill, C.D. 2020. Walking tall: a narrative critical family history of a grandmother's fight for new normal. *Genealogy* 4:58. DOI: 10.3390/genealogy4020058.

Tanner-Tremaine, P. n.d. *British 1820 Settlers to South Africa*. Available: https://1820settlers.com. [2025, January 17].

Theunissen, C.A. 2022. The effects of DNA test results on biological and family identities. *Genealogy* 6:17. DOI: 20.3390/genealogy6010017.

The National Archives, n.d. *Census records*. Available: https://www.nationalarchives.gov.uk/help-with-your-research/research-guides/census-records/ [2024, November 29]

Thomas, J., Helm, M.L., Helm, A.L and Barratt, N. 2007. *Genealogy online for dummies*. Chichester, West Sussex: John Wiley & Sons.

Tisdell, E.J. and Merriam, S.B. 2016. Chapter One: What is qualitative research? In *Qualitative Research*. United States: John Wiley & Sons. 3-21.

Tuffour, I. 2017. A critical overview of interpretative phenomenological analysis: a comtemporary qualitative research approach. *Journal of Healthcare Communication* 2(4):52. DOI: 10.4172/2472-1654.100093.

United Nations. 1948. *Universal Declaration of Human Rights*. Available: https://www.un.org/en/about-us/universal-declaration-of-human-rights. [2023, August 27].

Wallace, S.E., Gourna, E.G., Nikolova, V. and Sheehan, N.A. 2015. Family tree and ancestry inference: is there a need for a 'generational' consent? *BMC Medical Ethics* 16:87. DOI: 10.1186/s12910-015-0080-2.

Walters, P. 2019. *Ethical dilemmas in genealogy*. United Kingdom

Willever-Farr, H., Zach, L. and Forte, A. 2012. Tell me about my family: a study of cooperative research on Ancestry.com. *Proceedings of the 2012 iConference*. 7-10 February 2012. New York: Association for Computing Machinery. 303-310 DOI: 10.1145/2132176.2132215

Willever-Farr, H. and Forte, A. 2014. Family matters: control and conflict in online family history production. *CSCW 2014*. 15-19 February 2014, Baltimore: ACM. DOI:10.1145/2531602.253173

Wortzel, J.D., Wiebe, D.J., Elahi, S., Agawu, A., Barg, F.K. and Emmett, E.A. 2021. Ascertainment bias in a historic cohort study of residents in an asbestos manufacturing community*. International Journal of Environmental Research and Public Health* 18:2211. DOI: 10.3390/ijerph19052211.

Yakel, E. 2004. Seeking information, seeking connections, seeking meaning: genealogists and family historians. *Information Research* 10(1): 205. Available: http://files.eric.ed.gov/fulltext/EJ1082067.pdf. [2023, May 14].

Yakel, E. 2007. Digital curation. *OCLC Systems & Services: International Digital Library Perspectives*. 23(4):335-340. DOI: 10.1108/10650750710831466.

Yakel, E. and Torres, D.A. 2007. Genealogists as a "community of records". *The American Archivist* 70(1):93-113. Available: https://www.jstor.org/stable/40294451. [2023, March 1].

Yiotis, K. 2005. The open access initiative: a new paradigm for scholarly communications. *Information Technology and Libraries* 24:157-162. DOI: 10.6017/ital.v24i4.3378.

Zimmer, C. 2018. *She has her mother's laugh*. London: Picador.

Appendix A: Research Notes

A1. Research Hints

4 SP

Specificity of date and place: Civil registration indexes give me district of birth/ marriage and month it happened, whereas 1939 register gives exact address and date of birth (and occupation). Also, for 1939 register, information of living people redacted.

Record of baptism gives parish and county (district/parish/county: UK administrative geography can be confusing!), mother's address and occupation at time of birth – no father listed

Date of baptism vs date of birth

Description of baptism record gives name as Stephen Annie Porter, but document itself clearly separates name of son and mother.

Workhouse admission/discharge record at age of 3.

5 JW

1939 register updated – Porter crossed out and Hillard (surname of second husband) written in.

6 WL

Specificity of place names: Aston/Birmingham

Name of church in baptism record (and address where parents living)

World War 1 service record is that of his father –lists family members of serviceman

B/ham as abbreviation for Birmingham on above record

Surprised to see South African Estate/Death Record –first impression is that "Western Cape" is anachronistic here (High Court), but usage predates post-1994 province.

English Death Record gives his place of death as Tunbridge Wells, Kent, but he died in South Africa and his ashes were flown to England (no attached image to see details)

7 ET

Postal code in address for Death Index and district/subdistrict codes in Civil Registration Death Index

When adding parents to tree, Ancestry assumes father's surname to be that of child.

Ancestry gives predictions of parents based on other trees: some are nicknames and some have surnames in capital letters, depending on what is written in member tree.

8 ES – No hints

9 AP

Name as listed on records:

Ann – 1881 census

Annie – 1891, 1911 census, workhouse admission, Death Index

Anna – Poor Law District School District (but Annie on document – incorrect transcription)

4SP's baptism included with her records

Comment on census records and death index – document records age at the time (sometimes approximate), therefore birth year estimated.

10 HW

School admission: father listed as Will^{m} on document.

Probate calendar lists his death place as Torquay, not Croydon

Birthplace: 1891 census: Brompton; Birth Index: Medway…. Town vs District

1891 census: mother listed as Estella (actually father's second wife)

1939 register: updating of name of daughter.

11 CC

Birth: 2 March, but listing for 1939 register says 2 Feb. Transcription error because document itself says 2 March

Thought: would everyone look at document to check information?

Birth Index says Apr (indexed for second quarter – Apr-May-Jun)

Location information included on census: birth place, civil parish, residence place, registration district, sub registration district.

12 HL

Photo submitted by member: name given is Harry Richard

Select Marriage: birth place is London Warwick (sic) No image given

Civil Registration Index has his name as Richard Harry.

Harry on index, Henry on Ancestry description

Marriage Index & Banns: Henry Richard

His WW1 service record not listed?

13 AD

3 member-submitted photos

12HL's WW1 service record

1911 census: B'ham; Husband and son listed as Harry

Burial and cremation Index gives: d. 25 Aug 1968, bur. 3 Sep 1968 Greenwich London

Am I Wrong? YES – the death date I cited comes from a member tree!

14 CT

Geneanet Community Tree Index – user-made tree on another site (and has variant spelling)

15 EL – Only a tree hint

16 FS

1891 census: Registration district – St Saviour Southwark; Sub-registration district – St Mary Newington. Again, how do the districts work? How are they named?

Saints nor present on 1901 & 1911 census

Baptism record attached is that of son

17 LR

1891 census: Bermondsey transcribed as Dermonday

1871 census: same pattern (registration districts) as seen on 1891 (saints); ecclesiastical district/ parish on earlier census

Again, baptism record is that of son – she signed her full name as LEKR –is the K a previous marriage?

K also included on marriage to 16FS

18 JP

1851 census record – surname incorrectly transcribed, but correct on document. Also Chelmsford transcribed as Chensford. Date of birth is ~1824, on 1881 census is ~1831. What is correct date? Were there multiple JPs being conflated? Is the one in Islington (baptism record) 18JP or someone else?

Marriage record and baptism records with no images attached, cannot cross-check information. Baptism records appear to be for his son.

Spurious newspaper birth announcement from 1880

School admission record with residential address.

19 PS

Funeral paperwork submitted as member photo

Spelling: middle name Ley on Civil Birth Index, Leey on Marriage Index and Leigh on Death & Burial/Cremation Index.

Find-a-Grave Record has anachronistic location (Camberwell, London Borough of Southwark, Greater London.

Death Index gives birth year of 1836 (different to my records)

20 WW

1891 census – children as listed as how they are related to head of household, so natural assumption that wife of head is also mother of the children (she is not)

21ES present on 1881 census (where Guestling is also transcribed incorrectly as Guesbling)

20WW is described as widower on marriage to second wife and there is a marriage record to 21ES, where he is described as a bachelor

Two records are attached to him which are actually the baptisms of two of his children, one of which has 21ES's first name transcribed incorrectly.

21 ES

Burial record in 1888 – important considering 20WW remarried and census records featuring second wife.

22 CC

1851 census – variant spelling of name: Chandler instead of Chantler

Baptism record is that of his daughter.

23 EW – no reactions

24 RL

Two marriage records are those of his son

Where and when born:

Baptism: 1938, Winchcombe

1871, 1911 census: 1839, Gretton

1881, 1901 census: 1842, Winchcombe

1851 census: 1835, Winchcombe

1861 census: 1836, Gretton

1891 census: 1845, Gretton

25 JM

Two member-uploaded photographs

Name and when born:

Baptism: Julia Hannah, 1835

1861 census: Hannah, 1855

1871 census: Julia H, 1855

1881 census: Julianna, 1854

1891 census: Hannah, 1853

1901 census: Hannah, 1857

1911 census: Hannah, 1853

Also, 1891 census, Cleeve incorrectly transcribed as Cleis

26 WD

Member-submitted photo

1881 census: family recorded as Didzbury, not Didsbury. Mark on document lead to a transcription error.

1861 census: Name shortened to Wm; father's name to Richd

Death and Burial record is burial of child

Birth Index: West Riding of Yorkshire – again, administrative geography. Born in March, but listed as Apr, meaning registration in Apr-May-Jun quarter.

27 MG

1861 census – surname recorded as Griffiths

1871 census – surname incorrectly transcribed as Griffith, but correct on the document

Marriage in New Brunswick, Canada is that of his son Harry

28 HT

Geneanet Community Tree: surname spelled Terhoeven; place of marriage Cape Town, Western Cape, Afrique du Sud

29 DR – As for 28 HT

30 Unknown person

31 CL – No hints

32 ES – No hints

33 HD

Age discrepancy: 37 on 1881 census, 8 on 1861 census

34 Unknown person

35 LC

Married to JR, but both children surnamed KR and born before their marriage.

36 JP

Baptism records are those of his three children

Marriage record is that of his son

Pallot's Marriage Index lists 37CF as Flackston

37 CF

On parish register as Flaxman (no image)

On Geneanet Index as Flexman

38 JS

Regimental Register of Service: Description of physical appearance and height

39 MF

Two baptism records are of her son

Baptism record of a different Mary Ann Fowler (daughter of Willm and Elizth)

40 WW

Baptism record is of son

41 EW

Geneanet Record lists her husband as John. Spurious?

Two baptisms records for Hester (not Esther)

42/43 Unknown persons

44 JC

1851 census: as for 22CC

1861 census: written as Chantler, transcribed as Chandler

Spurious record for a John

One marriage marriage record is for his son 22CC

One marriage record for a Nathaniel Chandler (no image)

Two correct marriage records to 45MP

45 MP

Member photo of joint burial marker for 44JC and 45MP

Marriage record, 1851 and 1861 census records as seen for 44JC

Spurious burial record for a Mary Pratt from Norfolk

46 DW

Marriage record is that of daughter Eliza

47 RO

Geneanet Tree Record

Baptism record is that of son Daniel

48 WL

Member photo of coat of arms

Member photo of marriage certificate

Literacy: ‘X’ mark of bride

Also: Wm

Baptism and marriage records are of son Walter

Two hints of baptism records: one is son of Benjm and Mary, other is son of Daniel and Susannah

49 ES

Same member photo as for 48 WL

Baptism record transcribed as Stait (no image)

Marriage record says Staite

50 JM

1851 & 1871 census: surname recorded as Morriss

Marriage record is that of daughter Jane

One of the baptism records is of son Alfred

51 EH

1861 census: Covent Garden transcribed as Corent Garda

As for 50 JM, Morriss on 1851 census

52 RD

Abbreviations: 1871 census: Sheffd, Richd; 1861 census: Richd, Wm

Geneanet Community Tree

53 ME

Member photos of marriage and baptism records

Find a Grave record using modern administrative geography

54 HG

Three member photos by same member:

Marriage record (Griffett)

Baptism record (Griffiths)

1841 census record (Griffiths)

Griffiths records not same person as Griffett

Baptism record – parents not same as member-uploaded record

55CT dies 1871, 54HG remarries same year, remarries again 1874? Are they all the same person?

Death of death I recorded corresponds to HG who died in asylum in 1905 (admitted 1880)

1881 census record as Griffiths (different family?)

Do all these records belong to at least two different people?

55 CT

Marriage record to 54HG

1861 census as Griffiths

1871 & 1881 probably different family

56 CT

Only hint is an Ancestry tree

57 CV

Dutch Reform Church Register giving birth and baptism dates

City Directory: Mrs CMM Terhoeven

58-61 Unknown persons

62 AL

Baptism record is that of his daughter

63 CG

Same record as 62 AL

A2. Trees

4 SP

4 SP #4, 4 SP #5 and 4 SP # 9 copied from Geni.com

4 SP # 7 and 4 SP #16 both copied from me!

My father's tree (4 SP #11) includes speculative "Harry" as 4SP's father

4 SP # 3, 4 SP #4, 4 SP #5, 4 SP # 7 and 4 SP #9 have 4SP's siblings as half-siblings, probably correct, but probably only displayed as such because father not known

4 SP # 3 uses inconsistent geography: Lewisham, Greater London in 1972, but Lewisham, Kent in 1993

4 SP #4, 4 SP #5 and 4 SP # 9 adds "DNA" to 4SP's name to indicate DNA link, lists 5JW as Joan Hillard (Porter),

Pott4 SP #4, 4 SP #5 and 4 SP # 9: James 1831 Porter; Henry 1885 Wilson, William Albert 1991 Lawrence

4 SP #4, 4 SP #5 and 4 SP # 9: Ellen Adelaide Hopley-Lemmetjies (Copied from my father on Geni)

4 SP #4, 4 SP #5 and 4 SP # 9 list Estella West as 10HW's mother

4 SP #4, 4 SP #5 and 4 SP # 9: conflates 48WL with person of same name in New York, USA and New Brunswick, Canada and continues incorrectly many generations further back

4 SP #4, 4 SP #5 and 4 SP # 9: 18JP's children with 19PS and those from his previous marriage to Caroline Vickers are not all attributed to correct mother.

4 SP #4, 4 SP #5 and 4 SP # 9: seems to have lifted most of my tree (both sides) regardless of whether they are related to him

4 SP #10: TreeTag used!

5 JW

5 JW #1 gets death date and place wrong, confuses her with a Joan Cable who died 2001

5 JW #2 has 10 HW as Henry (Harry) Wilson and his mother as Estella Wilson

5 JW #3 has the same (copied from 5 JW #1, 5 JW #2 and 4 SP #7)

5 JW #3 includes Ellen Eliza Standen, but shows Estella as 10 HW's mother

5 JW #2 and 5 JW #4 both copied from my tree

6 WL

6 WL #1 makes extensive use of notations: (A/E15/19/Sa)(12steps)% or (A/9/1/11/15/19)(11steps)% and ranks (Sapper)

6 WL #1 also uses county names "South Yorkshire" and "West Midlands" before those counties existed

6 WL #2 copied from 6 WL #1 and uses the same notations and place names

Both have my deceased aunt (copied from my tree?) but not living aunts (information not available to them.

6 WL #3 has Birmingham as Blham and also only has deceased aunt

6 WL #6: Blham for Birmingham

6 WL #11: deceased aunt's tree, minimal detail

6 WL #12: also hers, minimal detail

6 WL #17: no sources, but has lifted information from somewhere about my family, such as father's death and information about 7 ET's side of family

6 WL #17 uses Western Cape for 1800's events, but also inconsistent in use.

6 WL #2: on summary, 13 AD's residence appears as Berkeley, California – is actually Berdesley Green, Bham

8 ES

As I discovered 8 ES to be my great grandfather through DNA evidence and he has acknowledged descendants who are most likely not aware of his paternity of 4 SP, I have removed names and surnames for privacy reasons.

8 ES #3: 8 ES's grandparents as JR & LK – her maiden name was C and her children both have K as part of their name. She married R after birth of children. K may be name of previous husband, but no record of marriage or death.

8 ES #1: also R & K

8 ES #12: same, but children triplicated on tree

8 ES #7: uses C as maiden name, but still assumes R as father.

8 ES #9: includes 8ES's first wife

8 ES #3 and 8 ES #7: marriage of 16 FS and 17 LR shows in summary as Newington, Oxfordshire.

9 AP

9 AP #1: 37 CF's surname as Flaxman

9 AP #9: Flackston

9 AP #6 and 9 AP #10: use Greater London as county name in 1905 (anachronistic)

9 AP #11: 37 CF as Flexman, her grandfather as Flaxman

9 AP #13: 19 PS as Prudance

9 AP # 17 and 9 AP #18: 9 AP's husband assumed to be Harold (trees of father's cousin)

10 HW

10 HW #3: Estella as mother, but Ellen Eliza listed as biological. Half-brother duplicated six times on tree

10 HW #4: Henry (Harry)

10 HW #7: Estella as mother

10 HW #7: brother of 10 HW appears twice – once as son of Ellen, once as son of Estella; sister shown as daughter of Ellen

10 HW #10: A thought – the descendants of Eileen Wilson are disproportionally represented because there are multiple family members who have taken to studying their family history, so that line will appear even when others are missing. Plus, users copy each other's' trees.

10 HW #12: some people have just dates entered, or just town with no county.

Also, I must bear in mind that for many of these trees, my family is just a side branch and of no particular interest

10 HW #16 & 10 HW #18-#30: Estella West's family is the main interest

10 HW #18-#30: multiple iteration of same tree – versioning, but numbering (including multiple date and AutoBackup) is messy

10 HW #34 and 10 HW #35: includes post codes in the places of birth and death.

11 CC

11 CC #2: Because I know there are users who have conflated different people with the same name (one English, one American) I sometimes find myself doubting a case of someone in England being connected to a person who was born/lived/died in the United States, despite not knowing anything about either of the people, because they are off my direct line and I have not studied their family history.

11 CC #3: many users might add siblings of ancestors, but not fill out the tree with their descendants, because they are not on their direct line. Also, for many people there are only names and dates, with no place information.

After looking at trees for so long, unless there is something unusual that stands out, there is often nothing worth commenting on, so I get the feeling that I am going through the motions and missing things.

11 CC #9: 11 CC and three siblings shown as children of 22 CC only and not 23 EW, unlike the other siblings

11 CC #11: parents of 22 CC given as James Chantler (b. 1794) and Mary Muggeridge; most other trees have James Chantler (b. 1802) and Mary Pratt (b. 1802). Confusion between two people of same name: other JC born in different town and county not associated with family. Mistakes elevate in older generations, so increasingly removed from real ancestor.

11 CC #14s: 11 CC duplicated on tree

12 HL

12 HL #1: 12 HL born Upton, Oxford, Maine USA, died Ottawa, Ontario, Canada – appears as such on summary, but correct information shown on page! No country information given, so website algorithms make assumptions?

12 HL #9: death in Ottawa on summary; 13AD 1915 residence in Mexico on summary, but in profile correctly as Berdesley Green, Birmingham

12 HL #15: marriage in Aston, Delaware, Pennsylvania (summary) – has lack of detail on page.

12 HL #14 copied from 6 WL #1 (rank and notation)

13 AD

13 AD #2 has one comment!

13 AD #3: has copied notation from 6 WL #1

13 AD #4: removed the notations, but place descriptions are the same as 6 WL #1 (also copied via 13 AD #3)

13 AD #5 has comments!

13 AD #7: has Griffett line their tree as Griffiths. I have not studied that part of my family enough to know who is right, them or me

6 WL #3: summary page weirdness for 13 AD: marriage in North Carolina, residence in California.

17 LR

17 LR #1: 17 LR and her brother duplicated

17 LR #4: comments!

18 JP

18 JP #1: has 18 JP's details, but name attached is Caroline Clasen; also conflated with another person of same name as 18 JP

18 JP #3: Has only 18 JP's first marriage; 37 CF as Flackston

18 JP #4: 9 AP's brother James immigrated to USA, along with half-brother

18 JP #5: focus is on family of wife of 9 AP's brother William

18 JP #7, 18 JP #8 and 18 JP #9 probably same person, creating new accounts (almost identically named)

18 JP #14: get the impression that uninteresting side branches have had minimal info included.

Many trees look much alike, even though other have not been cited as sources. Possible that many of them were not copied directly on the websites, but via notes written down.

A few trees are public, but have no publicly available family members, so functionally private

18 JP #32 conflating 18 JP with a different person of same name that was born in 1824 in Reading and was living in Berkshire in 1891.

19 PS

Looking at layout of trees: many include siblings of direct ancestors and do not fill out the tree unless of particular interest, such as trying to establish how related to DNA match

19 PS #5: profile photo of 19 PS is funeral paperwork – originally uploaded by another member and shows which other members saved it to their trees

19 PS #7: 19 PS listed as Prudence Ley Fowel Stephens

20 WW

20 WW #1: 20 WW married to Estella (no 21 ES) but has William as their child

20 WW #2: same names annotated with red squares or circles; Uses Tree Tags!

20 WW #3: Marriage and death of 20WW shows up in summary page as Greenwich, Priute, Utah!

20 WW #8: 20 WW's brother repeated on tree six times

20 WW #10: 40 WW used middle name or nickname Craddoch. Whether or not members use the name will affect what hints they receive and how they construct their trees

21 ES

On summary page, numerous trees list 21 ES 's birthplace as St Leonards, Tasmania, Australia

Correct on profile as Sussex, England

General comment about members using census, baptism or birth record as profile pic for individuals.

General comment: not valid to comment on accuracy of trees when the member has more information than I have or they have filled out the tree in places I have not

22 CC

22 CC #11: Comment!

22 CC #17: Father of 44 JC listed as James, not John and tree deviates from mine at earlier generations

22 CC #20 - #26: (same member as Rolf): has the same, but also has different mother for 22CC and different mother for 44JC; two different James Chantlers as husband of Mary Muggeridge and both Mary Muggeridge and 45 MP and wives of the one James Chantler.

22 CC #20 - #26: multiple people on the tree are repeated, sometimes multiple times; in some cases, where there is uncertainty about parentage, member has included both on the tree.

22 CC #27: multiple duplications of people on tree.

22 CC #32 and 22 CC #37: 22 CC as Chantler; 44 CC as Chandler; many first names in lower case

22 CC #36: A thought: some surnames written regularly, some written all in capitals, maybe because they are copied from another member?

22 CC #39: 22 CC as Chandler, other siblings Chandler or Chantler

22 CC #49 and 22 CC #52: Birth dates only, no death dates (part of tree unimportant for them?)

22 CC #51: People on tree who were born in UK, but died in USA – immigration and naturalisation records are cited. But also, people moving in other direction. Could be mistake, due to similarity of place names and no citations.

22 CC #58: Comments!

22 CC #59: duplication of people on tree, but some have same name as deceased sibling who died very young. But also, no citation.

22 CC #63 - #65: Tree Tags! (iterations of same tree)

22 CC #67: person named Emeriah (elsewhere as Emery)

22 CC #67: person in tree who emigrated from England to Canada – use profile pic that says Immigrated, but no comment or Tag

22 CC #72: Conflated 22 CC with American of same name – birth and parents is same as 22 CC, but later life and descendants is of different person. Copied from 22 CC #73, who also copied from a lot of others. Same mistake in 22 CC # 74 tree and 22 CC #25

23 EW

Choice of profile picture (general comment): if no photo or pre-photo era: coat of arms, gravestones, birth certificates, picture of home town

4 SP #10: marriage of 23 EW in Queens, Prince Edward Island, Canada in summary

23 EW #5: use of 'Not Known' when surname not known, instead of leaving blank.

23 EW #13: Tree Tags!

24 RL

24 RL #1: has included profession (innholder) in name of 24 RL's grandfather

24 RL #10: 25 JM as Juliet

25 JM

25 JM #1: has names of parents, supported by baptism documents (information I do not have). Could be correct, but that might be common name, even in a small village.

26 WD

26 WD #1: Anachronistic usage of Metropolitan borough of Calderdale/Kirklees, West Yorkshire for Ann Woodhead

26 WD #3: Anachronistic usage of Metropolitan borough of Sheffield, South Yorkshire for William Stocks

26 WD #4: Tree Tags!

26 WD #5: Anachronistic usage of West Midlands

26 WD #5: Names: Darwin / Derwent

26 WD #8: Wm for William

26 WD #8: Samuel/Lemuel Woodhead – uncertain transcription because of cursive writing

26 WD #10 and #11: Wm, Richd, Sheffd

26 WD #14: Farum/Fayrham (uses both spellings together)

26 WD #16: Wm – 1861 census as source

26 WD #20: anachronistic usage of West Midlands; names: Darwin/Darwinte;

26 WD #20: Comments!

26 WD #21: multiple duplicates of several people

26 WD #22 and 26 WD #24: both have Woodhead as Lemuel, not Samuel

26 WD #27: multiple duplications on tree

26 WD #27 and 26 WD #30: both have uncertainty and disagreement about parentage and marriages – is it really a complicated family history or a matter of confusing people of same name?

26 WD #6 and 26 WD #33: both use numbers instead of place names

27 MG – Only additional tree is private

28 HT

Only child of 28HT listed is the one who changed his name to Trehaeven

29 DR – No additional trees

30 Unknown person

31 CL

31 CL #1: assumes 15EL to be daughter of later husband

Appendix B: Other Tree Data Captured

Table B1 lists all the Ancestry.com member trees I looked at, the number of people on each tree, the time since the member who constructed the tree had last been active on Ancestry.com, and the number of other trees each member had constructed.

Tables B2-B28 list all trees on which a specific ancestor appeared, the number of records attached to that person on that tree, the sources that had been cited for their information, and the trees that the member who constructed the tree had cited as having copied from. A tree may appear more than once on a table if an ancestor had been duplicated on that tree.

The anonymised trees were named according to the earliest ancestor in the ahnentafel which appeared on that tree.

DP in Tables B2-B28 refers to my personal Ancestry.com tree, which some trees may have copied from

Private trees are shaded grey in the tables.

Table B1. All member trees studied

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
4 SP #1	15764	Sep 2005	< 1 day	
4 SP #2	102497	Oct 2007	< 1 day	
4 SP #3	1464	Aug 2021	> 1 year	
4 SP #4	400371	Jun 2020	< 1 day	
4 SP #5	395039	Jul 2022	2-6 days	1
4 SP #6	13240	Oct 2002	< 1 day	
4 SP #7	16083	May 2021	2-6 days	69
4 SP #8	2228	Nov 2016	< 1 day	
4 SP #9	422211	Jun 2020	< 1 day	1
4 SP #10	32494	Mar 2012	3-11 months	
4 SP #11	5	Sep 2000	> 1 year	
4 SP #12	71238	Aug 2006	> 1 year	
4 SP #13	71240	Aug 2006	> 1 year	
4 SP #14	71235	Aug 2006	> 1 year	
4 SP #15	37442	Aug 2006	> 1 year	
4 SP #16	52	Nov 2016	> 1 year	
4 SP #17	57852	Aug 2006	2-6 days	
5 JW #1	51281	Oct 2006	< 1 day	
5 JW #2	669	May 2021	> 1 year	
5 JW #3	31781	Apr 2015	1 day	
5 JW #4	39240	Aug 2013	2-4 weeks	
5 JW #5	15442	Nov 2006	< 1 day	
6 WL #1	68444	Sep 2010	3-11 months	
6 WL #2	31139	Jan 2016	< 1 day	
6 WL #3	32411	Oct 2006	< 1 day	
6 WL #4	4218	Aug 2009	1-2 weeks	
6 WL #5	2727	Aug 2007	< 1 day	
6 WL #6	797	Jan 2016	> 1 year	
6 WL #7	8555	Mar 2002	1-2 weeks	
6 WL #8	4939	Mar 2002	1-2 weeks	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
6 WL #9	4024	Mar 2002	1-2 weeks	
6 WL #10	287	May 2023	2-6 days	
6 WL #11	7	Dec 2019	> 1 year	
6 WL #12	4	Dec 2019	> 1 year	
6 WL #13	6306	May 2010	< 1 day	
6 WL #14	7743	Nov 2006	1 day	
6 WL #15	474	Jan 2012	< 1 day	
6 WL #16	357	Oct 2008	> 1 year	
6 WL #17	94845	Feb 2016	< 1 day	
6 WL #18	8130	Feb 2017	< 1 day	
8 ES #1	7950	Mar 2006	1-2 weeks	
8 ES #2	13707	Jan 2022	> 1 year	
8 ES #3	24373	Nov 2003	< 1 day	
8 ES #4	85	Aug 2009	2-6 days	
8 ES #5	21	Aug 2018	> 1 year	
8 ES #6	100080	Apr 2008	1 day	
8 ES #7	2010	Jun 2020	2-6 days	
8 ES #8	2508	Sep 2011	2-4 weeks	
8 ES #9	6650	Sep 2011	2-4 weeks	
8 ES #10	205	Jan 2020	> 1 year	
8 ES #11	1073	Nov 2004	1-2 weeks	
8 ES #12	491	Jan 2006	3-11 months	
9 AP #1	42439	March 2006	1 day	
9 AP #2	5720	Feb 2006	3-11 months	
9 AP #3	169	Mar 2001	1-3 months	
9 AP #4	349	Mar 2001	1-3 months	
9 AP #5	446	Mar 2001	1-3 months	
9 AP #6	193	Mar 2001	1-3 months	
9 AP #7	2231	Mar 2001	1-3 months	
9 AP #8	946	Jan 2006	< 1 day	11
9 AP #9	1579	Apr 2009	< 1 day	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
9 AP #10	10785	Jul 2009	2-4 weeks	
9 AP #11	10346	Mar 2001	< 1 day	91
9 AP #12	458	Jan 2010	> 1 year	
9 AP #13	2418	Jan 2005	< 1 day	
9 AP #14	198	Jul 2013	> 1 year	
9 AP #15	3357	Jun 2006	2-4 weeks	3
9 AP #16	69708	Mar 2007	> 1 year	3
9 AP #17	7	Jan 2010	> 1 year	3
9 AP #18	12	Jan 2010	> 1 year	
10 HW #1	555	Jul 2003	2-6 days	27
10 HW #2	54	Jul 2003	2-6 days	
10 HW #3	6635	Feb 2014	< 1 day	
10 HW #4	11937	Jul 2006	< 1 day	28
10 HW #5	8164	Jul 2006	< 1 day	
10 HW #6	208981	Aug 2004	< 1 day	
10 HW #7	15346	Apr 2020	< 1 day	
10 HW #9	2462	Jul 2006	< 1 day	
10 HW #10	15588	Jun 2013	1 day	
10 HW #11	7053	Jul 2018	3-11 months	
10 HW #12	4049	Jul 2002	< 1 day	19
10 HW #13	2156	Aug 2008	2-4 weeks	
10 HW #14	2129	Apr 2015	> 1 year	2
10 HW #15	7219	Nov 2007	> 1 year	
10 HW #16	1490	May 2010	> 1 year	
10 HW #17	2928	Jun 2012	3-11 months	
10 HW #18	3944	Jan 2006	< 1 day	
10 HW #19	2909	Jan 2006	< 1 day	
10 HW #20	4251	Jan 2006	< 1 day	
10 HW #21	3721	Jan 2006	< 1 day	
10 HW #22	3685	Jan 2006	< 1 day	
10 HW #23	4415	Jan 2006	< 1 day	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
10 HW #24	2809	Jan 2006	< 1 day	
10 HW #25	5024	Jan 2006	< 1 day	
10 HW #26	2721	Jan 2006	< 1 day	
10 HW #27	4004	Jan 2006	< 1 day	
10 HW #28	6093	Jan 2006	< 1 day	
10 HW #29	5868	Jan 2006	< 1 day	
10 HW #30	1389	Jan 2006	< 1 day	
10 HW #31	197	Jul 2011	> 1 year	
10 HW #32	80033	Sep 2006	> 1 year	
10 HW #33	13751	Nov 2004	> 1 year	
10 HW #34	79953	Jul 2005	2-6 days	
10 HW #35	29884	Jul 2005	2-6 days	
10 HW #36	35475	Jul 2005	2-6 days	
10 HW #37	35810	Oct 2012	> 1 year	
10 HW #38	34437	Oct 2012	> 1 year	
10 HW #39	34436	Aug 2020	> 1 year	
10 HW #40	6899	Mar 2011	> 1 year	
11 CC #1	10861	Apr 2007	> 1 year	
11 CC #2	19780	Jul 2012	> 1 year	
11 CC #3	3706	Aug 2015	> 1 year	
11 CC #4	558	Aug 2010	> 1 year	
11 CC #5	8438	Dec 2007	1-3 months	
11 CC #6	3834	Jan 2006	< 1 day	
11 CC #7	7597	Nov 2012	1-2 weeks	
11 CC #8	377	Jul 2014	> 1 year	
11 CC #9	15710	Nov 2002	< 1 day	
11 CC #10	11390	Dec 2000	1-2 weeks	
11 CC #11	10789	Mar 2023	< 1 day	
11 CC #12	15150	Apr 2016	< 1 day	
11 CC #13	775	Jun 2019	> 1 year	
11 CC #14	5301	Sep 2019	< 1 day	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
11 CC #15	1912	Aug 2010	1-2 weeks	
12 HL #1	104	Nov 2021	3-11 months	
12 HL #2	157	Sep 2022	> 1 year	
12 HL #3	147	Feb 2014	3-11 months	
12 HL #4	147	Feb 2014	3-11 months	
12 HL #5	70631	Sep 2021	1-2 weeks	
12 HL #6	111922	Sep 2021	1-2 weeks	
12 HL #7	277801	Feb 2015	1-3 months	
12 HL #8	3523	Aug 2007	< 1 day	
12 HL #9	110	Nov 2017	> 1 year	
12 HL #10	218	Jan 2007	1-2 weeks	
12 HL #11	105	Nov 2006	3-11 months	
12 HL #12	12259	Jan 2006	> 1 year	
12 HL #13	40	Nov 2016	1 day	
12 HL #14	52709	Mar 2010	< 1 day	
12 HL #15	41996	Jan 2006	2-6 days	
12 HL #16	10828	Aug 2008	2-6 days	
13 AD #1	9377	<2000	< 1 day	
13 AD #2	818	Feb 2013	1-2 weeks	
13 AD #3	1130	Nov 2006	> 1 year	
13 AD #4	1754	Jan 2023	< 1 day	
13 AD #5	13943	Sep 2008	2-4 weeks	
13 AD #6	1812	June 2012	< 1 day	
13 AD #7	24298	Apr 2012	2-6 days	
13 AD #8	2378	Dec 2000	2-6 days	
13 AD #9	1285	Jul 2008	> 1 year	
13 AD #10	48	Feb 2017	3-11 months	
13 AD #11	718	May 2002	1 day	
13 AD #12	5141	Jan 2012	1-2 weeks	
14 CT #1	32305	Jun 2010	< 1 day	
15 EL #1	8	Feb 2022	3-11 months	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
16 FS #1	21360	Jan 2004	< 1 day	
16 FS #2	6312	Feb 2015	< 1 day	8
16 FS #3	5085	Dec 2010	1-3 months	
16 FS #4	4394	Dec 2010	1-3 months	
17 LR #1	916	Jun 2020	< 1 day	
17 LR #2	916	Jun 2020	< 1 day	
17 LR #3	42404	Aug 2006	3-11 months	
17 LR #4	5089	Jul 2009	3-11 months	
17 LR #5	36010	Feb 2009	1-3 months	
17 LR #6	111873	Jun 2002	< 1 day	
17 LR #7	125792	Mar 2002	1-2 weeks	
17 LR #8	2177	Aug 2012	> 1 year	
17 LR #9	854	Nov 2010	3-11 months	
18 JP #1	142	Feb 2022	3-11 months	
18 JP #2	210808	Apr 2005	< 1 day	12
18 JP #3	596	Jan 2018	3-11 months	2
18 JP #4	446	Mar 2001	1-2 weeks	
18 JP #5	2939	Sep 2005	2-6 days	
18 JP #6	809	Feb 2007	3-11 months	6
18 JP #7	20283	Oct 2009	3-11 months	
18 JP #8	23197	Jan 2012	1-3 months	
18 JP #9	20291	Sep 2014	3-11 months	
18 JP #10	116344	Feb 2014	< 1 day	301
18 JP #11	3034	Jul 2010	3-11 months	
18 JP #12	9926	Jul 2008	2-6 days	
18 JP #13	4800	Apr 2013	2-6 days	
18 JP #14	3227	Jan 2006	2-6 days	
18 JP #15	3625	Jan 2006	2-6 days	
18 JP #16	6747	Jan 2006	2-6 days	
18 JP #17	2724	Jan 2008	> 1 year	
18 JP #18	6465	Jul 2001	> 1 year	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
18 JP #19	7964	Jul 2001	> 1 year	
18 JP #20	10230	Jun 2003	1 day	18
18 JP #21	5602	Dec 2000	1-2 weeks	
18 JP #22	5571	Dec 2000	1-2 weeks	
18 JP #23	629	Sep 2009	2-6 days	
18 JP #24	579	Sep 2009	2-6 days	
18 JP #25	7625	Jun 2003	2-6 days	
18 JP #26	3642	Sep 2007	2-6 days	
18 JP #27	1241	Dec 2015	< 1 day	
18 JP #28	7168	Jul 2005	1-3 months	
18 JP #29	966	Jan 2012	3-11 months	
18 JP #30	1623	Jun 2003	2-6 days	
18 JP #31	11877	Dec 2000	3-11 months	
18 JP #32	758	Sep 2009	1-2 weeks	
19 PS #1	1682	Oct 2006	< 1 day	
19 PS #2	727	Mar 2011	< 1 day	
19 PS #3	37725	Sep 2010	< 1 day	
19 PS #4	47211	Oct 2010	< 1 day	
19 PS #5	3371	Apr 2011	> 1 year	
19 PS #6	9576	Dec 2022	3-11 months	
19 PS #7	364	Jun 2005	1-2 weeks	
19 PS #8	5955	Jul 2013	1-3 months	
19 PS #9	409	Feb 2006	2-6 days	
19 PS #10	2185	Apr 2021	2-6 days	
19 PS #11	1223	Feb 2016	3-11 months	
20 WW #1	283	Aug 2023	1-3 months	
20 WW #2	4311	Apr 2019	3-11 months	
20 WW #3	9127	Jul 2005	2-6 days	
20 WW #4	5104	Aug 2012	< 1 day	
20 WW #5	3691	Nov 2005	2-6 days	
20 WW #6	26115	Mar 2002	< 1 day	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
20 WW #7	1940	Jun 2009	2-6 days	
20 WW #8	189	Jan 2019	3-11 months	
20 WW #9	7540	Apr 2020	3-11 months	
20 WW #10	22759	Mar 2007	< 1 day	
20 WW #11	19563	Apr 2019	< 1 day	
20 WW #12	1384	Feb 2010	3-11 months	
20 WW #13	1081	Nov 2004	< 1 day	
20 WW #14	5652	Jul 2016	< 1 day	
20 WW #15	197	Sep 2010	3-11 months	
20 WW #16	800	Sep 2008	> 1 year	
21 ES #1	4064	Jul 2015	< 1 day	
21 ES #2	269	Apr 2012	1-3 months	
21 ES #3	48579	Jul 2007	2-6 days	
22 CC #1	34589	Jun 2018	< 1 day	
22 CC #3	32162	Nov 2019	< 1 day	
22 CC #4	12530	Jul 2005	< 1 day	
22 CC #5	2064	Mar 2009	1-2 weeks	
22 CC #6	975	Jan 2006	< 1 day	
22 CC #7	1798	Sep 2002	< 1 day	
22 CC #8	1090	May 2006	> 1 year	
22 CC #9	1091	Jul 2009	> 1 year	
22 CC #10	1051	Jul 2009	> 1 year	
22 CC #11	8480	Sep 2011	< 1 day	9
22 CC #12	3892	Jun 2007	1 day	11
22 CC #13	626	Jun 2011	3-11 months	1
22 CC #14	169685	Jan 2013	< 1 day	7
22 CC #15	10297	Mar 2006	< 1 day	
22 CC #16	101	Nov 2004	2-6 days	16
22 CC #17	60891	Aug 2014	3-11 months	3
22 CC #18	6652	Nov 2004	1 day	
22 CC #18	89	Aug 2018	> 1 year	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
22 CC #19	89	Aug 2018	> 1 year	
22 CC #20	91083	Aug 2014	3-11 months	
22 CC #21	91083	Aug 2014	3-11 months	
22 CC #22	91083	Aug 2014	3-11 months	
22 CC #23	91083	Aug 2014	3-11 months	
22 CC #24	91083	Aug 2014	3-11 months	
22 CC #25	91083	Aug 2014	3-11 months	
22 CC #26	91083	Aug 2014	3-11 months	
22 CC #27	3155	Oct 2004	2-6 days	17
22 CC #28	15794	Jan 2017	1-2 weeks	
22 CC #29	21315	Mar 2018	< 1 day	
22 CC #30	195	Dec 2021	3-11 months	
22 CC #31	204	Nov 2004	2-6 days	
22 CC #32	663	Nov 2013	< 1 day	7
22 CC #33	123639	Jan 2006	3-11 months	
22 CC #34	151	Mar 2005	< 1 day	
22 CC #35	143953	Nov 2008	< 1 day	
22 CC #36	153	Sep 2010	> 1 year	
22 CC #37	5384	Oct 2006	> 1 year	3
22 CC #38	293	Dec 2020	> 1 year	
22 CC #39	676	May 2020	1-2 weeks	
22 CC #40	16334	Aug 2001	1-3 months	
22 CC #41	50	Sep 2015	> 1 year	
22 CC #42	154	Sep 2022	> 1 year	
22 CC #43	574	Jun 2007	1 day	
22 CC #44	378	Aug 2018	> 1 year	
22 CC #45	176270	Jun 2004	> 1 year	
22 CC #46	775	Jul 2018	1-2 weeks	
22 CC #47	151	Oct 2012	2-6 days	
22 CC #48	783	Oct 2010	< 1 day	
22 CC #49	489	Jun 2011	3-11 months	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
22 CC #50	828	May 2014	> 1 year	
22 CC #51	850	Sep 2007	> 1 year	
22 CC #52	818	Jul 2002	1-3 months	
22 CC #53	399	May 2014	1-3 months	
22 CC #54	7284	Mar 2023	1-2 weeks	
22 CC #55	2570	Dec 2020	1-3 months	
22 CC #56	7053	Jul 2018	1-2 weeks	
22 CC #57	793	Jan 2008	3-11 months	
22 CC #58	5109	Apr 2003	1 day	
22 CC #59	4602	Jan 2002	< 1 day	
22 CC #60	6968	Sep 2008	< 1 day	
22 CC #61	3720	Sep 2014	3-11 months	
22 CC #62	10191	Nov 2010	Today	
22 CC #63	9413	Feb 2023	> 1 year	
22 CC #64	9413	Feb 2023	> 1 year	
22 CC #65	9420	Feb 2023	> 1 year	
22 CC #66	9413	Oct 2019	> 1 year	
22 CC #67	12514	Jan 2017	3-11 months	
22 CC #68	3323	Jan 2017	3-11 months	
22 CC #69	5487	Jul 2005	< 1 day	
22 CC #70	2004	May 2003	1-2 weeks	
22 CC #71	132	Jul 2017	3-11 months	
22 CC #72	12413	Dec 2001	3-11 months	
22 CC #73	4085	Jul 2007	2-6 days	
22 CC #74	136	Feb 2018	3-11 months	
23 EW #1	31942	Aug 2015	1-3 months	
23 EW #2	1319	Aug 2007	3-11 months	4
23 EW #3	546	Aug 2007	3-11 months	
23 EW #4	32266	Feb 2006	< 1 day	
23 EW #5	2219	Jun 2004	< 1 day	
23 EW #6	2219	Jun 2004	< 1 day	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
23 EW #7	9887	Oct 2008	1-2 weeks	
23 EW #8	312	Aug 2007	3-11 months	
23 EW #9	3920	Feb 2018	1 day	
23 EW #10	3920	Feb 2018	1 day	
23 EW #11	2444	Mar 2013	1-3 months	
23 EW #12	3389	Jan 2011	2-6 days	
23 EW #13	11921	Oct 2006	< 1 day	
23 EW #14	1380	Apr 2010	3-11 months	
23 EW #15	23858	Jul 2009	1-3 months	
23 EW #16	496126	Jan 2014	> 1 year	
23 EW #17	3367	Mar 2016	1-3 months	
23 EW #18	3984	Sept 2007	3-11 months	
23 EW #19	607	Dec 2013	> 1 year	
23 EW #20	607	Dec 2013	> 1 year	
23 EW #21	147	Jul 2012	> 1 year	
23 EW #22	761	Jun 2004	< 1 day	
23 EW #23	1142	Feb 2014	1 day	
23 EW #24	1561	Feb 2005	< 1 day	
23 EW #25	37704	Mar 2005	1-3 months	
23 EW #26	37680	Mar 2005	1-3 months	
23 EW #27	7361	Nov 2007	1-2 weeks	
24 RL #1	6557	Oct 2003	< 1 day	
24 RL #2	42300	Sep 2010	< 1 day	
24 RL #3	9567	Mar 2009	1-3 months	
24 RL #4	63	Sep 2010	1-2 weeks	
24 RL #5	5153	Feb 2019	1-2 weeks	
24 RL #6	3161	Jun 2006	< 1 day	
24 RL #7	18169	Jun 2014	3-11 months	
24 RL #8	4149	Mar 2014	< 1 day	
24 RL #9	52	May 2003	over a year	
24 RL #10	76	Nov 2016	< 1 day	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
24 RL #11	25	Sep 2007	1 day	
25 JM #1	5788	May 2005	< 1 day	
25 JM #2	1275	May 2009	1-2 weeks	
25 JM #3	165	Oct 2012	3-11 months	
25 JM #4	4902	Jan 2013	3-11 months	
26 WD #1	71940	Feb 2007	1-2 weeks	
26 WD #2	3692	Jun 2011	1-3 months	
26 WD #3	3692	Jun 2011	1-3 months	
26 WD #3	6725	Aug 2007	2-6 days	
26 WD #4	2900	Oct 2007	< 1 day	
26 WD #5	140527	Dec 2000	< 1 day	
26 WD #6	849	Nov 2005	3-11 months	
26 WD #7	576	Jul 2007	1-3 months	
26 WD #8	5339	Nov 2005	2-6 days	
26 WD #9	225	Jan 2009	3-11 months	
26 WD #10	1460	Dec 2014	3-11 months	
26 WD #11	1642	Dec 2014	3-11 months	
26 WD #12	20821	Apr 2004	> 1 year	
26 WD #13	466	Jan 2017	> 1 year	
26 WD #14	4590	Aug 2003	< 1 day	
26 WD #15	23824	Jun 2006	2-6 days	
26 WD #16	2383	May 2010	1-3 months	
26 WD #17	184	Jul 2016	1-3 months	
26 WD #18	47727	Apr 2006	< 1 day	
26 WD #19	2958	Dec 2003	1-2 weeks	
26 WD #20	528	Jul 2010	3-11 months	
26 WD #21	1496	Sep 2006	3-11 months	
26 WD #22	265	Dec 2018	3-11 months	
26 WD #23	13260	Jun 2007	< 1 day	
26 WD #24	2381	Jan 2002	< 1 day	
26 WD #25	101547	before 2000	< 1 day	

Tree Name	Number of people on tree	Date member joined Ancestry.com	Time since member last active on Ancestry.com	Number of other trees
26 WD #26	751	Sep 2011		
26 WD #27	865	Dec 2014	> 1 year	
26 WD #28	6572	Apr 2006	2-6 days	
26 WD #29	361856	Feb 2011	< 1 day	
26 WD #30	601	Nov 2004	2-6 days	
26 WD #31	37924	Feb 2006	2-6 days	6
26 WD #32	13087	before 2000	2-6 days	
26 WD #33	54748	Aug 2002	2-6 days	
26 WD #34	446	Mar 2013	3-11 months	
26 WD #35	446	Mar 2013	3-11 months	
26 WD #36	207	Nov 2018	1-3 months	
26 WD #37	21895	Feb 2005	1-3 months	
26 WD #38	22100	May 2008	1-3 months	
26 WD #39	20750	May 2008	1-3 months	
26 WD #40	21346	May 2008	1-3 months	
26 WD #41	20350	May 2008	1-3 months	
26 WD #42	20785	May 2008	1-3 months	
26 WD #43	20278	May 2008	1-3 months	
26 WD #44	17408	May 2008	1-3 months	
26 WD #45	19764	May 2008	1-3 months	
26 WD #46	19810	May 2008	1-3 months	
26 WD #47	20245	May 2008	1-3 months	
26 WD #48	20779	May 2008	1-3 months	
26 WD #49	20882	May 2008	1-3 months	
26 WD #50	84	Sep 2011	1-3 months	
26 WD #51	13483	Aug 2002	< 1 day	
26 WD #52	115138	Jan 2001	2-6 days	
26 WD #53	115138	Jan 2001	2-6 days	
27 MG #1	3814	Mar 2018	1-2 weeks	
28 HT #1	59333	Mar 2005	< 1 day	
31 CL #1	28613	May 2005	< 1 day	

Table B2. Trees on which 4 SP appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #1	7	13	
4 SP #2	8	10	
4 SP #3	9	9	
4 SP #4	0	2	
4 SP #5	0	2	
4 SP #6	3	4	
4 SP #7	0	1	DP
4 SP #8	0	1	
4 SP #9	0	2	
4 SP #10	0	0	
4 SP #11	0	0	
4 SP #12	0	0	
4 SP #13	0	0	
4 SP #14	0	0	
4 SP #15	0	0	
4 SP #16	0	1	DP
4 SP #17	0	0	

Table B3. Trees on which 5 JW appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #1	7	10	
4 SP #2	4	8	
4 SP #3	3	3	
4 SP #4	0	3	
4 SP #5	0	3	
4 SP #6	3	5	
4 SP #7	0	1	DP
4 SP #8	0	1	
4 SP #9	0	3	
4 SP #10	0	0	
4 SP #11	0	0	
4 SP #12	0	0	
4 SP #13	0	0	
4 SP #16	0	1	DP
5 JW #1	3	3	
5 JW #2	0	3	DP; 4 SP #16; 5 JW #1
5 JW #3	0	3	
5 JW #4	0	1	DP
5 JW #5	1	1	

Table B4. Trees on which 6 WL appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #7	0	1	DP
4 SP #8	4	9	
4 SP #9	0	1	
6 WL #1	4	6	DP; 6 WL #6
6 WL #2	3	4	6 WL #1
6 WL #3	4	4	
6 WL #4	3	3	
6 WL #5	2	3	
6 WL #6	7	7	
6 WL #7	0	0	
6 WL #8	0	1	
6 WL #9	0	1	
6 WL #10	3	7	
6 WL #11	0	0	
6 WL #12	2	2	
6 WL #13	3	4	
6 WL #14	1	2	
6 WL #15	1	1	
6 WL #16	0	1	
6 WL #17	0	0	
6 WL #18	0	0	

Table B5. Trees on which 7 ET appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #7	0	1	DP
4 SP #8	1	2	
4 SP #9	0	1	
6 WL #1	2	3	DP
6 WL #2	1	2	6 WL #1
6 WL #3	1	1	
6 WL #4	2	2	
6 WL #5	0	0	
6 WL #7	0	0	
6 WL #11	0	0	
6 WL #17	0	0	

Table B6. Trees on which 8 ES appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
8 ES #1	7	11	
8 ES #2	0	2	8 ES #12
8 ES #3	5	6	8 ES #2, 8 ES #5, 8 ES #6
8 ES #4	0	3	
8 ES #5	0	0	8 ES #5
8 ES #6	0	1	8 ES #12
8 ES #7	1	2	8 ES #12
8 ES #8	4	4	
8 ES #9	4	4	
8 ES #10	1	1	
8 ES #11	1	1	
8 ES #12	5	5	
8 ES #12	2	2	

Table B7. Trees on which 9 AP appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #1	11	19	
4 SP #3	11	11	-
4 SP #7	0	1	DP
4 SP #8	0	1	
4 SP #11	0	0	
9 AP #1	8	8	
9 AP #2	7	7	
9 AP #3	2	3	9 AP #6
9 AP #4	2	2	
9 AP #5	2	2	
9 AP #6	1	5	9 AP #3, 9 AP #4, 9 AP #8
9 AP #7	1	1	
9 AP #8	2	4	9 AP #3, 9 AP #6
9 AP #9	1	1	
9 AP #10	1	2	9 AP #4
9 AP #11	1	2	9 AP #1
9 AP #12	1	1	
9 AP #13	1	1	
9 AP #13	1	1	
9 AP #14	0	6	
9 AP #15	0	1	9 AP #17
9 AP #16	0	1	
9 AP #17	0	0	
9 AP #18	0	0	

Table B8. Trees on which 10 HW appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #2	9	20	
4 SP #3	1	1	
4 SP #4	0	2	
4 SP #5	0	2	
4 SP #6	1	4	
4 SP #7	0	1	DP
4 SP #9	0	2	
4 SP #10	0	0	
4 SP #12	0	0	
4 SP #13	0	0	
4 SP #14	0	0	
4 SP #15	0	0	
4 SP #16	0	2	DP
5 JW #1	2	2	
5 JW #2	5	11	DP, 4 SP #10, 4 SP #16, 5 JW #1, 10 HW #4
5 JW #3	5	10	4 SP #7, 5 JW #1, 5 JW #2, 10 HW #3, 10 HW #4
5 JW #4	0	1	DP
5 JW #5	1	1	
10 HW #1	6	8	
10 HW #2	1	1	
10 HW #3	2	7	DP, 5 JW #4, 10 HW #410 HW #12, 10 HW #14,
10 HW #4	5	7	
10 HW #5	2	2	
10 HW #6	5	5	
10 HW #7	2	4	10 HW #3, 10 HW #12
10 HW #8	0	0	
10 HW #9	2	2	
10 HW #10	1	1	
10 HW #11	0	0	
10 HW #12	2	2	

Tree Name	Number of Records	Number of Sources	Trees Cited
10 HW #13	1	1	
10 HW #14	2	3	
10 HW #15	0	2	
10 HW #16	2	2	
10 HW #17	0	1	
10 HW #18	0	0	
10 HW #19	0	0	
10 HW #20	0	0	
10 HW #21	0	0	
10 HW #22	0	0	
10 HW #23	0	0	
10 HW #24	0	0	
10 HW #25	0	0	
10 HW #26	0	0	
10 HW #27	0	0	
10 HW #28	0	0	
10 HW #29	0	0	
10 HW #30	0	0	
10 HW #31	0	1	10 HW #30
10 HW #32	0	0	
10 HW #33	0	2	10 HW #16
10 HW #34	0	0	
10 HW #35	0	0	
10 HW #36	0	0	
10 HW #37	0	0	
10 HW #38	0	0	
10 HW #39	0	1	
10 HW #40	0	1	10 HW #12

Table B9. Trees on which 11 CC appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #2	7	16	
4 SP #3	7	7	
4 SP #4	0	2	
4 SP #5	0	2	
4 SP #6	6	9	
4 SP #7	0	1	DP
4 SP #9	0	2	
4 SP #10	2	2	
4 SP #12	0	0	
4 SP #13	0	0	
4 SP #14	0	0	
4 SP #15	0	0	
4 SP #16	0	2	DP
4 SP #17	0	0	
5 JW #1	6	6	
5 JW #2	8	14	DP, 4 SP #10, 4 SP #16, 5 JW #1, 10 HW #4
5 JW #3	6	14	4 SP #3, 4 SP #7, 5 JW #1, 5 JW #2, 10 HW #3, 10 HW #4, 10 HW #10, 11 CC #2
5 JW #4	2	5	DP, 5 JW #1, 10 HW #10
5 JW #5	7	7	
10 HW #1	7	8	
10 HW #2	1	1	
10 HW #3	1	4	DP, 5 JW #4, 10 HW #4
10 HW #4	7	7	
10 HW #5	7	7	
10 HW #6	7	7	
10 HW #9	4	4	
10 HW #10	6	6	
10 HW #11	0	0	
11 CC #1	3	4	

Tree Name	Number of Records	Number of Sources	Trees Cited
11 CC #2	2	4	10 HW #9, 10 HW #10
11 CC #3	2	3	Deleted
11 CC #4	2	2	
11 CC #5	2	2	
11 CC #6	2	2	
11 CC #7	0	1	Deleted
11 CC #8	1	1	
11 CC #9	2	2	
11 CC #10	0	0	
11 CC #11	0	4	
11 CC #12	0	1	10 HW #10
11 CC #13	0	0	
11 CC #14	0	1	
11 CC #15	0	0	

Table B10. Trees on which 12 HL appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #7	0	1	DP
4 SP #8	11	20	
4 SP #9	0	1	
6 WL #1	0	1	DP
6 WL #1	4	8	DP, 6 WL #6, 12 HL #11
6 WL #2	4	6	6 WL #1, 12 HL #9
6 WL #3	8	8	
6 WL #4	11	11	
6 WL #5	3	3	
6 WL #6	7	7	
6 WL #7	0	0	
6 WL #8	0	1	
6 WL #9	0	2	
6 WL #10	2	3	
6 WL #10	12	18	
6 WL #13	10	12	
6 WL #14	6	8	
6 WL #15	3	3	
6 WL #16	4	5	
6 WL #18	0	1	
12 HL #1	0	0	
12 HL #2	0	0	
12 HL #3	2	2	
12 HL #4	0	1	
12 HL #5	0	0	
12 HL #6	0	0	
12 HL #7	1	1	
12 HL #8	5	9	

Tree Name	Number of Records	Number of Sources	Trees Cited
12 HL #9	3	3	
12 HL #10	1	1	
12 HL #11	0	0	
12 HL #12	1	2	12 HL #11
12 HL #13	0	1	6 WL #2
12 HL #14	0	3	6 WL #1, 12 HL #1, 12 HL #11
12 HL #15	0	0	
12 HL #16	0	0	

Table B11. Trees on which 13 AD appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #7	0	1	DP
4 SP #8	13	23	
4 SP #9	0	1	
6 WL #1	5	10	DP; 6 WL #6, 6 WL #15, 12 HL #11
6 WL #2	9	11	6 WL #1, 12 HL #9
6 WL #3	5	5	
6 WL #4	7	7	
6 WL #5	4	4	
6 WL #6	11	11	
6 WL #7	1	1	
6 WL #8	0	1	
6 WL #9	0	2	
6 WL #10	13	14	
6 WL #10	7	7	
6 WL #13	8	11	
6 WL #14	5	6	
6 WL #15	3	4	DP
6 WL #16	1	2	
6 WL #18	0	0	
12 HL #1	0	0	
12 HL #9	3	3	
12 HL #10	0	0	
12 HL #11	0	0	
12 HL #12	4	6	
12 HL #13	0	3	6 WL #2, 6 WL #6, 13 AD #4
12 HL #14	0	3	6 WL #1, 12 HL #1, 12 HL #11
12 HL #15	0	0	
12 HL #16	2	3	

Tree Name	Number of Records	Number of Sources	Trees Cited
13 AD #1	3	3	
13 AD #2	3	3	
13 AD #3	1	3	6 WL #1, 13 AD #2
13 AD #4	1	4	6 WL #1, 13 AD #2, 13 AD #3
13 AD #5	1	3	6 WL #15, 13 AD #2
13 AD #6	1	1	
13 AD #7	1	1	
13 AD #8	1	1	
13 AD #9	1	1	
13 AD #10	0	5	12 HL #12, 13 AD #1, 13 AD #5, 13 AD #7
13 AD #11	0	1	
13 AD #12	0	0	

Table B12. Trees on which 14 CT appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #9	0	1	
6 WL #1	0	1	DP
6 WL #17	0	0	
14 CT #1	0	0	

Table B13. Trees on which 15 EL appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #9	0	1	
6 WL #1	0	1	DP
6 WL #17	0	0	
14 CT #1	0	0	
15 EL #1	0	0	

Table B14. Trees on which 16 FS appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
8 ES #1	6	9	
8 ES #2	0	2	
8 ES #3	5	6	8 ES #12
8 ES #7	3	4	8 ES #12
8 ES #8	0	0	
8 ES #8	0	0	
8 ES #9	0	0	
8 ES #12	3	3	
8 ES #12	2	2	
8 ES #12	2	2	
8 ES #12	1	1	
16 FS #1	3	3	
16 FS #2	2	2	
16 FS #3	1	1	
16 FS #4	1	1	

Table B15. Trees on which 17 LR appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
8 ES #1	8	13	
8 ES #2	0	3	
8 ES #3	7	9	8 ES #12, 17 LR #5
8 ES #7	7	10	8 ES #3, 8 ES #12, 17 LR #1
8 ES #12	7	7	
8 ES #12	5	5	
8 ES #12	5	5	
8 ES #12	4	4	
8 ES #12	3	3	
8 ES #12	2	2	
8 ES #12	2	2	
8 ES #12	2	2	
16 FS #1	7	11	
17 LR #1	1	4	17 LR #5, 17 LR #8
17 LR #1	1	1	
17 LR #2	0	3	
17 LR #3	1	2	
17 LR #4	2	2	
17 LR #5	0	3	
17 LR #6	0	0	
17 LR #7	2	2	
17 LR #8	0	1	

Table B16. Trees on which 18 JP appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #1	13	28	
4 SP #1	1	1	
4 SP #3	1	1	
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #7	0	1	DP
4 SP #9	0	1	
9 AP #1	23	24	9 AP #3
9 AP #2	12	12	
9 AP #9	4	4	
9 AP #3	11	15	9 AP #6, 18 JP #4
9 AP #3	1	1	
9 AP #4	2	4	
9 AP #6	3	11	9 AP #3, 9 AP #6
9 AP #6	1	1	
9 AP #7	3	3	
9 AP #8	5	8	9 AP #3, 9 AP #6
9 AP #10	1	2	9 AP #4
9 AP #11	7	8	9 AP #1
9 AP #12	1	2	9 AP #3
9 AP #13	1	1	
9 AP #14	0	6	
18 JP #1	9	9	
18 JP #2	11	11	
18 JP #3	0	8	DP, 9 AP #1, 9 AP #6, 9 AP #8, 9 AP #12, 18 JP #10, 18 JP #12, 18 JP #23
18 JP #4	2	2	
18 JP #5	2	3	DP
18 JP #6	2	6	
18 JP #7	2	2	

Tree Name	Number of Records	Number of Sources	Trees Cited
18 JP #8	2	2	
18 JP #9	2	2	
18 JP #10	0	6	9 AP #1, 18 JP #14, 18 JP #20, 18 JP #29, 18 JP #32
18 JP #11	2	2	
18 JP #12	0	1	9 AP #1
18 JP #13	0	0	
18 JP #14	0	0	
18 JP #15	0	0	
18 JP #16	0	0	
18 JP #17	0	0	
18 JP #18	0	0	
18 JP #19	0	0	
18 JP #20	0	0	
18 JP #21	0	0	
18 JP #22	0	0	
18 JP #23	0	1	9 AP #3
18 JP #24	0	1	
18 JP #25	0	0	
18 JP #26	0	0	
18 JP #27	1	1	
18 JP #28	0	0	
18 JP #29	0	1	
18 JP #30	1	2	
18 JP #31	0	0	
18 JP #32	1	2	

Table B17. Trees on which 19 PS appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #1	10	18	
4 SP #3	3	3	
4 SP #7	0	1	DP
9 AP #1	11	13	9 AP #3, 9 AP #6
9 AP #2	11	11	
9 AP #3	11	18	9 AP #5, 9 AP #6, 9 AP #8
9 AP #4	5	7	
9 AP #5	6	6	
9 AP #6	9	17	9 AP #3, 9 AP #4, 9 AP #5, 9 AP #8
9 AP #7	3	3	
9 AP #8	9	12	9 AP #3, 9 AP #6
9 AP #9	1	1	
9 AP #10	1	2	9 AP #4
9 AP #11	2	3	9 AP #1
9 AP #12	1	2	9 AP #3
9 AP #13	3	3	
9 AP #14	1	7	
9 AP #16	0	1	
18 JP #1	5	5	
18 JP #2	0	0	
18 JP #5	10	11	DP
18 JP #6	8	12	
18 JP #30	8	15	
18 JP #32	1	1	
19 PS #1	5	5	
19 PS #2	1	3	9 AP #3, 9 AP #6
19 PS #3	1	2	19 PS #2
19 PS #4	1	2	9 AP #3
19 PS #5	1	1	19 PS #2
19 PS #6	1	1	

Tree Name	Number of Records	Number of Sources	Trees Cited
19 PS #7	0	0	
19 PS #8	0	2	19 PS #3
19 PS #9	1	2	19 PS #5
19 PS #10	0	1	9 AP #3
19 PS #11	2	2	

Table B18. Trees on which 20 WW appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #2	14	20	
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #9	0	1	
5 JW #2	1	2	DP
5 JW #3	3	11	5 JW #2, 10 HW #3, 10 HW #7, 10 HW #13, 10 HW #14, 20 WW #2, 20 WW #6, 20 WW #8
10 HW #1	7	9	
10 HW #1	9	10	
10 HW #3	4	9	DP, 10 HW #12, 10 HW #14, 20 WW #6, 20 WW #8
10 HW #7	13	15	10 HW #3, 10 HW #12
10 HW #12	4	4	
10 HW #13	4	4	
10 HW #14	5	6	
10 HW #15	0	1	
10 HW #16	1	1	
10 HW #17	0	1	
10 HW #18	0	0	
10 HW #19	0	0	
10 HW #20	0	0	
10 HW #21	0	0	
10 HW #22	0	0	
10 HW #23	0	0	
10 HW #24	0	0	
10 HW #25	0	0	
10 HW #26	0	0	
10 HW #27	0	0	
10 HW #28	0	0	
10 HW #29	0	0	
10 HW #30	0	0	

Tree Name	Number of Records	Number of Sources	Trees Cited
10 HW #31	0	1	10 HW #30
10 HW #32	2	2	
10 HW #33	1	2	10 HW #31
10 HW #33	0	1	10 HW #16
10 HW #34	1	1	
10 HW #35	0	0	
10 HW #36	0	0	
10 HW #37	0	0	
10 HW #38	0	0	
10 HW #39	0	1	
10 HW #40	0	1	10 HW #12
20 WW #1	1	1	
20 WW #2	1	6	DP, 10 HW #13, 20 WW #6, 20 WW #16
20 WW #3	0	0	
20 WW #4	2	2	
20 WW #5	2	3	
20 WW #6	2	2	
20 WW #7	1	3	
20 WW #8	1	8	DP, 10 HW #12, 10 HW #37, 20 WW #3, 20 WW #10, 20 WW #12, 20 WW #16
20 WW #9	2	2	
20 WW #10	1	1	
20 WW #11	1	1	
20 WW #12	1	1	
20 WW #13	0	1	20 WW #6
20 WW #14	0	1	10 HW #13
20 WW #15	0	1	
20 WW #16	0	1	20 WW #3

Table B19. Trees on which 21 ES appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #2	9	15	
4 SP #6	0	1	
5 JW #3	0	1	10 HW #3
10 HW #1	2	3	
10 HW #3	1	2	DP
10 HW #7	10	12	10 HW #3, 21 ES #2
21 ES #1	0	0	
21 ES #2	0	0	
21 ES #3	3	3	

Table B20. Trees on which 22 CC appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #3	9	9	
4 SP #4	0	0	
4 SP #5	0	0	
4 SP #6	5	7	
4 SP #9	0	0	
4 SP #10	3	3	
4 SP #12	0	2	
4 SP #13	0	2	
4 SP #14	0	2	
4 SP #15	0	1	
4 SP #17	1	4	
5 JW #1	9	9	
5 JW #2	2	5	DP , 5 JW #1
5 JW #3	8	13	4 SP #3, 5 JW #2, 10 HW #4, 10 HW #10, 11 CC #2
5 JW #4	2	6	DP, 5 JW #1, 10 HW #10, 22 CC #13
5 JW #5	7	7	
10 HW #1	26	27	
10 HW #3	1	1	
10 HW #4	11	11	
10 HW #5	10	10	
10 HW #6	9	9	
10 HW #9	6	6	
10 HW #10	41	42	11 CC #5
11 CC #1	8	17	
11 CC #2	12	14	10 HW #9, 10 HW #10
11 CC #3	8	21	22 CC #4, 22 CC #9, 22 CC #11, 22 CC #15, 22 CC #20, 22 CC #31, 22 CC #36, 22 CC #37, 22 CC #51, 22 CC #52, 22 CC #53, 22 CC #57
11 CC #4	5	5	
11 CC #5	11	11	

Tree Name	Number of Records	Number of Sources	Trees Cited
11 CC #6	6	7	
11 CC #7	4	14	5 JW #1, 10 HW #4, 11 CC #5, 11 CC #13, 22 CC #9, 22 CC #13, 22 CC #45, 22 CC #58
11 CC #8	1	1	
11 CC #9	7	7	
11 CC #10	0	0	
11 CC #11	0	16	
11 CC #12	0	2	10 HW #10, 11 CC #5
11 CC #13	0	0	
11 CC #15	1	1	
22 CC #1	10	10	
22 CC #2	2	2	
22 CC #3	8	8	
22 CC #4	0	0	
22 CC #5	3	4	10 HW #10
22 CC #6	1	1	
22 CC #7	1	2	
22 CC #8	1	3	5 JW #1, 10 HW #4
22 CC #9	1	4	22 CC #20
22 CC #10	1	2	
22 CC #11	1	2	Deleted
22 CC #12	1	2	5 JW #4
22 CC #13	1	14	5 JW #1, 10 HW #10, 11 CC #3, 11 CC #5, 22 CC #11, 22 CC #27, 22 CC #31, 22 CC #45, 22 CC #49, 22 CC #57
22 CC #14	1	2	11 CC #13
22 CC #15	1	1	
22 CC #16	1	1	
22 CC #17	1	2	22 CC #68
22 CC #18	1	1	
22 CC #18	0	0	
22 CC #19	0	0	

Tree Name	Number of Records	Number of Sources	Trees Cited
22 CC #20	1	2	22 CC #27
22 CC #21	0	2	22 CC #11, 22 CC #57
22 CC #22	0	1	Deleted
22 CC #23	1	2	22 CC #73
22 CC #24	0	1	22 CC #73
22 CC #25	0	10	22 CC #9, 22 CC #15, 22 CC #36, 22 CC #68, Other #1
22 CC #26	0	1	22 CC #73
22 CC #27	2	3	11 CC #9
22 CC #28	1	1	
22 CC #29	3	3	
22 CC #30	0	1	Deleted
22 CC #31	0	1	Deleted
22 CC #32	0	1	5 JW #1
22 CC #33	1	1	
22 CC #34	0	0	
22 CC #35	0	1	10 HW #10
22 CC #36	0	2	22 CC #15; 22 CC #57
22 CC #37	0	1	Deleted
22 CC #38	0	3	5 JW #1, 10 HW #4
22 CC #39	0	1	Deleted
22 CC #40	1	1	
22 CC #41	0	2	22 CC #11, 22 CC #51
22 CC #42	0	1	22 CC #13
22 CC #43	0	1	11 CC #5
22 CC #44	0	3	5 JW #1, 11 CC #5
22 CC #45	0	4	5 JW #1, 11 CC #5
22 CC #46	0	1	
22 CC #47	0	1	
22 CC #48	0	1	Deleted
22 CC #49	1	1	
22 CC #50	0	1	

Tree Name	Number of Records	Number of Sources	Trees Cited
22 CC #51	0	1	22 CC #68
22 CC #52	0	1	22 CC #9
22 CC #53	0	1	22 CC #52
22 CC #54	0	1	22 CC #51
22 CC #55	0	1	
22 CC #56	0	0	
22 CC #57	0	1	
22 CC #58	0	1	11 CC #13
22 CC #59	0	0	
22 CC #60	0	0	
22 CC #61	0	1	
22 CC #62	0	1	
22 CC #63	0	1	
22 CC #64	0	1	
22 CC #65	0	1	
22 CC #66	0	1	
22 CC #67	0	1	
22 CC #68	0	1	22 CC #57
22 CC #69	0	1	
22 CC #70	0	0	
22 CC #71	3	3	-
22 CC #72	12	21	22 CC #20, 22 CC #36, 22 CC #73, Other #2, Other #3, Other #4, Other #5, Other #6, Other #7
22 CC #73	6	18	22 CC #9, 22 CC #36, 22 CC #72, Other #5, Other #6, Other #8, Other #9, Other #10, Other #11, Other #12, Other #13
22 CC #74	0	0	

Table B21. Trees on which 23 EW appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #3	17	17	
4 SP #4	0	0	
4 SP #5	0	0	
4 SP #6	3	4	
4 SP #9	0	0	
4 SP #10	7	7	
4 SP #12	0	1	
4 SP #13	0	1	
4 SP #14	0	1	
4 SP #15	0	1	
4 SP #17	0	0	
5 JW #1	14	14	
5 JW #2	2	5	DP, 5 JW #1
5 JW #3	16	23	4 SP #, 5 JW #2, 10 HW #4, 10 HW #9, 10 HW #10, 11 CC #2, 23 EW #1
5 JW #4	2	4	5 JW #1, 10 HW #10
5 JW #5	5	5	
10 HW #1	32	32	
10 HW #3	1	1	
10 HW #4	12	13	
10 HW #5	17	17	
10 HW #6	9	9	
10 HW #9	10	10	
10 HW #10	38	41	10 HW #9, 11 CC #5, 23 EW #21
10 HW #11	0	0	
11 CC #1	8	12	
11 CC #2	17	20	10 HW #9, 10 HW #10, 23 EW #22
11 CC #3	7	8	22 CC #4
11 CC #4	10	10	
11 CC #5	9	9	

Tree Name	Number of Records	Number of Sources	Trees Cited
11 CC #6	4	4	
11 CC #7	2	3	Deleted
11 CC #8	1	1	
11 CC #9	4	6	23 EW #5, 23 EW #7
11 CC #10	0	0	
11 CC #11	0	11	
11 CC #12	0	1	10 HW #10
11 CC #14	0	1	10 HW #4
11 CC #15	0	0	
22 CC #1	3	3	
22 CC #2	1	1	
22 CC #3	17	17	
22 CC #4	0	0	
22 CC #5	3	4	10 HW #10
22 CC #6	0	0	
22 CC #18	1	1	
22 CC #18	1	1	
22 CC #29	1	1	
22 CC #33	1	1	
22 CC #34	0	0	
22 CC #35	0	2	10 HW #9, 10 HW #10
22 CC #40	1	1	
22 CC #60	0	0	
22 CC #67	0	2	
22 CC #69	0	0	
22 CC #70	0	0	
23 EW #1	3	16	DP, 4 SP #3, 4 SP #10, 4 SP #12, 10 HW #4, 10 HW #9, 10 HW #10, 11 CC #2, 11 CC #14, 23 EW #2, 23 EW #5, 23 EW #21, 23 EW #24
23 EW #2	3	4	10 HW #9
23 EW #3	2	5	4 SP #3, 23 EW #2, 23 EW #5

Tree Name	Number of Records	Number of Sources	Trees Cited
23 EW #4	2	2	
23 EW #5	1	2	10 HW #10
23 EW #6	0	5	
23 EW #7	1	1	
23 EW #8	1	1	
23 EW #9	1	1	
23 EW #10	1	1	
23 EW #11	1	1	
23 EW #12	1	1	
23 EW #13	1	1	
23 EW #14	1	1	
23 EW #15	1	1	
23 EW #16	0	3	
23 EW #17	1	1	
23 EW #18	0	7	
23 EW #19	0	1	
23 EW #20	0	1	
23 EW #21	0	1	11 CC #9
23 EW #22	0	0	
23 EW #23	0	1	
23 EW #24	0	1	11 CC #2
23 EW #25	0	0	
23 EW #26	0	0	
23 EW #27	0	1	

Table B22. Trees on which 24 RL appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #8	11	19	
6 WL #1	0	1	DP
6 WL #3	10	10	
6 WL #4	10	10	
6 WL #5	5	5	
6 WL #5	1	1	
6 WL #13	9	10	
6 WL #14	10	13	
6 WL #16	12	13	
6 WL #18	0	0	
12 HL #1	0	0	
12 HL #2	2	2	
12 HL #3	6	7	12 HL #10
12 HL #5	0	0	
12 HL #6	0	0	
12 HL #7	1	1	
12 HL #10	2	3	12 HL #11
12 HL #11	0	0	
24 RL #1	5	5	
24 RL #2	5	6	12 HL #7
24 RL #3	3	3	
24 RL #4	0	1	24 RL #1
24 RL #5	0	1	
24 RL #6	0	1	24 RL #1
24 RL #7	0	1	12 HL #3
24 RL #8	1	1	
24 RL #9	0	0	
24 RL #10	0	0	
24 RL #11	1	1	

Table B23. Trees on which 25 JM appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #8	10	20	
4 SP #9	0	1	
6 WL #1	0	1	DP
6 WL #3	6	6	
6 WL #4	8	8	
6 WL #5	0	0	
6 WL #10	5	10	
6 WL #10	1	1	
6 WL #13	7	8	
6 WL #14	7	9	
6 WL #16	9	10	
6 WL #18	0	1	
12 HL #1	1	1	
12 HL #2	3	3	
12 HL #3	5	6	12 HL #10
12 HL #5	0	0	
12 HL #6	0	0	
12 HL #7	1	1	
12 HL #8	8	10	
12 HL #10	2	2	
12 HL #11	1	1	
24 RL #2	4	9	12 HL #3, 12 HL #7, 12 HL #10, 12 HL #11, 24 RL #7
24 RL #7	0	1	12 HL #3
24 RL #9	0	0	
24 RL #11	0	0	
25 JM #1	3	3	
25 JM #2	2	2	
25 JM #3	0	2	DP, 25 JM #1

Tree Name	Number of Records	Number of Sources	Trees Cited
25 JM #4	1	1	

Table B24. Trees on which 26 WD appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #7	0	3	DP, 13 AD #3, 26 WD #3
4 SP #8	12	23	
4 SP #9	0	1	
6 WL #1	5	8	DP, 6 WL #6, 6 WL #15
6 WL #2	3	3	
6 WL #5	2	2	
6 WL #6	13	13	
6 WL #6	3	3	
6 WL #6	1	1	
6 WL #7	2	2	
6 WL #8	0	4	
6 WL #9	0	2	
6 WL #10	15	17	
6 WL #13	8	8	
6 WL #14	0	1	
6 WL #15	5	5	
6 WL #16	0	1	
6 WL #18	0	0	
12 HL #1	0	0	
12 HL #10	2	2	
12 HL #11	0	0	
12 HL #12	5	8	6 WL #15, 12 HL #11, 26 WD #6
12 HL #13	1	4	6 WL #2, 6 WL #6, 13 AD #4
12 HL #15	0	0	
12 HL #16	6	7	
13 AD #1	9	9	
13 AD #2	5	6	

Tree Name	Number of Records	Number of Sources	Trees Cited
13 AD #3	10	22	6 WL #1, 13 AD #2, 13 AD #4, 26 WD #2, 26 WD #9, 26 WD #14, 26 WD #31, 26 WD #33, 26 WD #34, 26 WD #35
13 AD #4	13	19	6 WL #1, 6 WL #15, 13 AD #2, 13 AD #3, 26 WD #25, 26 WD #33
13 AD #5	10	14	6 WL #15, 12 HL #10, 13 AD #2, 26 WD #6
13 AD #6	1	1	
13 AD #7	13	13	
13 AD #8	2	8	
13 AD #9	1	1	
13 AD #10	0	6	12 HL #12, 13 AD #1, 13 AD #5, 13 AD #7, 26 WD #3
13 AD #11	0	1	
13 AD #12	0	0	
26 WD #1	12	12	
26 WD #2	4	5	Deleted
26 WD #3	2	2	
26 WD #3	11	11	
26 WD #4	2	2	
26 WD #5	0	0	
26 WD #6	4	5	Deleted
26 WD #7	2	2	
26 WD #8	2	2	
26 WD #9	1	5	12 HL #10, 12 HL #11, 26 WD #14
26 WD #10	1	2	
26 WD #11	2	2	
26 WD #12	1	2	
26 WD #13	1	8	
26 WD #14	0	0	
26 WD #15	1	1	
26 WD #16	1	3	13 AD #1, 26 WD #11
26 WD #17	1	1	

Tree Name	Number of Records	Number of Sources	Trees Cited
26 WD #18	1	1	
26 WD #19	1	1	
26 WD #20	0	0	
26 WD #21	1	1	
26 WD #22	1	1	
26 WD #23	1	1	
26 WD #24	0	2	6 WL #15, 13 AD #4
26 WD #25	0	6	26 WD #3, 26 WD #14, 26 WD #31, 26 WD #33, 26 WD #34
26 WD #26	0	1	
26 WD #27	0	11	13 AD #3, 26 WD #3, 26 WD #9, 26 WD #15, 26 WD #31, 26 WD #33, 26 WD #34, 26 WD #35
26 WD #28	0	4	
26 WD #29	0	1	26 WD #3
26 WD #30	0	0	
26 WD #31	1	1	
26 WD #32	0	1	
26 WD #33	0	2	26 WD #9
26 WD #34	0	0	
26 WD #35	0	0	
26 WD #36	0	1	26 WD #16
26 WD #37	0	0	
26 WD #38	1	1	
26 WD #39	1	1	
26 WD #40	1	1	
26 WD #41	1	1	
26 WD #42	1	1	
26 WD #43	1	1	
26 WD #44	1	1	
26 WD #45	1	1	
26 WD #46	1	1	

Tree Name	Number of Records	Number of Sources	Trees Cited
26 WD #47	1	1	
26 WD #48	1	1	
26 WD #49	1	1	
26 WD #50	0	0	
26 WD #51	0	0	
26 WD #52	0	0	
26 WD #53	0	0	

Table B25. Trees on which 27 MG appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #7	0	1	DP
4 SP #8	13	22	
4 SP #9	0	1	
6 WL #1	7	11	6 WL #6, 6 WL #15, 12 HL #12
6 WL #2	1	1	
6 WL #5	1	1	
6 WL #6	13	13	
6 WL #6	8	8	
6 WL #7	1	1	
6 WL #8	0	2	
6 WL #9	0	1	
6 WL #10	16	18	
6 WL #13	2	4	
6 WL #14	0	1	
6 WL #15	3	3	
6 WL #16	0	1	
6 WL #18	0	0	
12 HL #1	0	0	
12 HL #10	1	1	
12 HL #11	0	0	
12 HL #12	11	14	6 WL #15, 12 HL #11, 26 WD #6
12 HL #13	1	4	6 WL #2, 6 WL #6, 13 AD #4
12 HL #15	0	0	
12 HL #16	5	6	
13 AD #1	11	11	
13 AD #2	5	6	
13 AD #3	5	8	6 WL #1, 13 AD #2, 26 WD #2
13 AD #4	14	18	6 WL #1, 12 HL #12, 13 AD #2, 13 AD #3

Tree Name	Number of Records	Number of Sources	Trees Cited
13 AD #5	10	13	6 WL #15, 13 AD #2, 26 WD #6
13 AD #6	1	1	
13 AD #7	15	15	
13 AD #8	1	1	
13 AD #9	1	1	
13 AD #10	0	6	12 HL #12, 13 AD #1, 13 AD #5, 13 AD #7, 26 WD #2
13 AD #12	0	0	
26 WD #1	8	8	
26 WD #2	2	2	
26 WD #3	8	8	
26 WD #4	1	1	
26 WD #5	0	0	
26 WD #6	7	7	
26 WD #14	0	0	
26 WD #20	0	0	
26 WD #23	1	1	
26 WD #32	0	1	
26 WD #37	0	0	
26 WD #51	0	0	
26 WD #52	0	0	
26 WD #53	0	0	
27 MG #1	1	1	

Table B26. Trees on which 28 HT appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #9	0	1	
6 WL #17	1	1	
14 CT #1	0	0	
28 HT #1	0	1	Deleted

Table B27. Trees on which 29 DR appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
4 SP #4	0	1	
4 SP #5	0	1	
4 SP #9	0	1	
6 WL #17	0	0	
14 CT #1	0	0	
28 HT #1	0	1	Deleted

Table B28. Trees on which 31 CL appeared

Tree Name	Number of Records	Number of Sources	Trees Cited
14 CT #1	0	0	
31 CL #1	0	0	

Appendix C: Ethical Clearance Letter

Department of Knowledge & Information Stewardship
University of Cape Town
Upper Campus

Private Bag X1, RONDEBOSCH, 7701 South Africa
Level 5 Hiangsanani, Chancellor Oppenheimer Library
Tel: +27 (0) 21 650 4546
E-mail: dkis@uct.ac.za
Website: www.dkis.uct.ac.za

10 April 2024
Ref: UCT DKIS2024-03-02
Ethics clearance valid for **12 months** from date of issue

Donovan Porter (PRTDON002)
Department of Knowledge and Information Stewardship
Chancellor Oppenheimer Library
University of Cape Town

Ethics approval for Master's research

Dear Donovan Porter

I am pleased to inform you that ethical clearance has been granted by the Ethics Review Committee of the Department of Knowledge and Information Stewardship (DKIS), on behalf of the Humanities Faculty of the University of Cape Town, for you to proceed with collecting data for your Master's study entitled: *Constructing a family tree in an online space as an act of digital curation.*

As a next step, please ensure that you obtain approval from the relevant online platform to collect, use and/or analyse any data hosted on the platform, if required by the platform's Terms and Conditions.

We wish you well with your data collection and the completion of your research.

Yours faithfully,

Mr Richard Higgs
Chair: Department (DKIS) Research Ethics Committee