

# Continguts

<b>1</b>	<b>Introducció al projecte</b>	<b>5</b>
1.1	Introducció . . . . .	5
1.2	Motivació i context . . . . .	6
1.3	Objectius . . . . .	7
1.4	Planificació . . . . .	7
1.4.1	Planificació Febrer del 2014 - Juliol 2014 . . . . .	8
1.4.2	Planificació Febrer del 2016 - Setembre 2016 . . . . .	8
<b>2</b>	<b>La genealogia</b>	<b>13</b>
2.1	Què és la genealogia? . . . . .	13
2.2	El paper de la genealogia en el transcurs de la història . . . . .	14
2.3	Les lleis reguladores . . . . .	15
2.4	Els codis ètics en la genealogia . . . . .	16
2.5	El procés de recerca genealògica . . . . .	17
2.6	Conclusió . . . . .	18
<b>3</b>	<b>L'organització FamilySearch</b>	<b>19</b>
3.1	Què és FamilySearch? . . . . .	19
3.2	La història de FamilySearch . . . . .	20
3.3	L'església mormona i la família . . . . .	21
3.4	Serveis per organitzacions amb arxius genealògics . . . . .	23
3.4.1	Captura d'imatges . . . . .	23
3.4.2	Conversió de formats digital . . . . .	24
3.4.3	Indexació en línea . . . . .	24
3.4.4	Accés en línea . . . . .	24

3.4.5	Preservació dels registres i fitxers físics . . . . .	24
3.4.6	Conclusions sobre els serveis professionals . . . . .	25
3.5	Serveis per particulars i aficionats a la genealogia . . . . .	25
3.5.1	Eines genealògiques . . . . .	26
3.6	Recerca genealògica . . . . .	29
3.6.1	Cerca de registres . . . . .	29
3.6.2	Cerca de genealogies . . . . .	30
3.6.3	Cerca per catàleg . . . . .	31
3.6.4	Cerca en llibres d'història genealògica . . . . .	31
3.6.5	Wiki de FamilySearch . . . . .	31
3.7	Projectes d'indexació . . . . .	31
3.8	Conclusió sobre les eines per particulars . . . . .	32
<b>4</b>	<b>Introducció a l'API de FamilySearch</b>	<b>33</b>
4.1	El portal de desenvolupadors . . . . .	33
4.2	L'arquitectura de l'Interfície de Programació d'Aplicacions (API) . .	34
4.2.1	Què és una API? . . . . .	34
4.2.2	L'arquitectura REST . . . . .	34
4.3	Formats de dades utilitzats pel Sistema . . . . .	36
4.3.1	El format de dades GEDCOM i GEDCOM X . . . . .	37
4.3.2	Format de dades FamilySearch . . . . .	38
4.3.3	Format de dades Atom (o Atòmic) . . . . .	39
4.3.4	Codificacions dels formats de dades . . . . .	39
4.4	Evolució temporal de l'API . . . . .	41
4.4.1	Evolucions al llarg del temps . . . . .	41
4.4.2	Tree Foundation . . . . .	41
4.5	Serveis extres oferts per l'API . . . . .	43
4.5.1	Caching . . . . .	44
4.5.2	Throttling . . . . .	44
4.5.3	Sincronització . . . . .	44
4.5.4	Internacionalització . . . . .	45
4.5.5	Transferència de registres en grans quantitats . . . . .	45
4.6	Eines de desenvolupament . . . . .	45

<i>CONTINGUTS</i>	3
4.6.1 Els SDK de FamilySearch . . . . .	46
4.6.2 Les aplicacions d'exemple . . . . .	47
4.6.3 Altres eines intressants . . . . .	47
4.7 Procés de certificació . . . . .	47
4.8 Diferents entorns de treball . . . . .	48
<b>5 Valoració final sobre la potencialitat de l'API</b>	<b>49</b>
5.1 Introducció . . . . .	49
5.2 Distribució geogràfica de les dades . . . . .	49
5.3 Dades contamporànies . . . . .	50
5.4 Recursos i funcionalitats . . . . .	51
5.5 Naturalesa de l'API . . . . .	51
5.6 Utilització de l'API en el marc d'un PFC . . . . .	52
5.7 Conclusió . . . . .	52
<b>6 Llista de propostes de projecte</b>	<b>55</b>
6.1 Introducció . . . . .	55
6.2 Comparació sobre la popularitat de noms . . . . .	55
6.3 Portal de cerca localitzat al Català . . . . .	56
6.4 Geolocalització d'un cognom en diferents nivells . . . . .	57
6.5 La recerca genealògica i l'heràldica . . . . .	57
6.6 Projectes d'indexació . . . . .	58
6.7 La història de l'església mormona a través de FamilySearch . . . . .	58
6.8 Diversitat geogràfica d'un cognom: Els nostres avantpassats . . . . .	59
6.9 Les col·leccions de dades de FamilySearch . . . . .	59
6.10 FamilySearch i la segona guerra mundial: Natalitat i Defuncions . . . . .	60
6.11 FamilySearch i la segona guerra mundial: Increment en els casaments . . . . .	60
6.12 FamilySearch i la segona guerra mundial: La llista de Schindler . . . . .	61
6.13 La gran recessió o altres esdeveniments històrics . . . . .	62
6.14 Comparacions amb els amics o seguidors de Facebook i Twitter . . . . .	62
6.15 Comparació de dades genealògiques reals amb FamilySearch . . . . .	63
6.16 Estudi de profunditat dels arbres familiars . . . . .	63
6.17 Algoritme de marcatge de duplicats . . . . .	64



## Secció 1

# Introducció al projecte

### 1.1 Introducció

Aquest projecte neix de les conversacions amb Enric Mayol mentre explorava diferents opcions sobre quin projecte final de carrera realitzar. L'Enric em va introduir l'organització de FamilySearch i l'existència de la seva API, encarregada de gestionar l'accés a les dades d'índole genealògic.

FamilySearch és una organització sense ànim de lucre destinada a connectar famílies a través de generacions. La seva visió com a col·lectiu és el d'ajudar a les persones a crear un vincle amb els seus avantpassats, com a eina per poder comprendre millor qui són, crear un sentiment de família i teixir el pont entre passat i futur.

Com s'ha esmentat, les dades emmagatzemades per l'organització i accessibles a través de l'API són principalment de caràcter genealògic. En concret, es disposa d'una col·lecció de persones de les quals se'n coneix informació personal, esdeveniments rellevants en el transcurs de la seva vida, com podrien ser per exemple dades sobre el seu naixement i les seves relacions amb altres persones, en altres paraules, el seu arbre genealògic.

El projecte gira entorn aquesta API i ha estat dividit en tres grans blocs o seccions.

Per començar, realitzar un estudi profund de l'API. Això significa comprendre quines són les petites peces d'informació realment disponibles i com estan relacionades entre elles.

En segon lloc, i com un dels tres blocs principals, utilitzant el coneixement adquirit durant l'estudi de l'API així com les oportunitats i limitacions imposades per la plataforma conegudes durant la implementació dels exemples, plantejar un conjunt de propostes de projecte que puguin servir a futurs estudiants com a suport i inspiració.

Finalment, l'últim bloc del projecte consisteix en implementar una aplicació que

interactuï amb l'API de FamilySearch a través de diferents exemples. L'objectiu dels exemples és el de facilitar l'observació i comprensió del potencial de l'API, exposar-ne la informació emmagatzemada i oferir idees sobre com encarar-ne l'explotació.

## 1.2 Motivació i context

Durant el transcurs de la carrera són moltes i diverses les vessants de la informàtica que ens van ser introduïdes. D'aquestes, sempre vaig sentir més afinitat per aquelles que requerien allunyar-se un pèl dels detalls més tècnics i s'enfocaven en un exercici d'abstracció i conceptualització centrat en la comprensió d'un problema quotidià, el disseny d'una solució i finalment, la seva execució.

Entenc doncs, que donada la situació, no és d'estranyar que sentís una empatia més elevada per camps com la mineria de dades o l'enginyeria del software que no pas l'algorítmica o els sistemes operatius. Al mateix temps, sempre m'he considerat una persona dispersa a qui li agrada conèixer una mica de molts temes diferents i per tot això, buscava un projecte que em permetés trencar en certa forma amb aquests aspectes més tècnics de la informàtica i explorar un camp desconegut al qual es poguessin aplicar els coneixements adquirits durant la carrera.

El projecte proposat per l'Enric complia doncs, en bona mesura, amb tot allò que jo buscava. La genealogia representava un camp que desconeixia per complet i el projecte en si era un full de ruta obert. Només la realització de la primera part d'aquest ens permetria comprendre com de profundes i completes podrien ser les propostes que el projecte originaria o quins serien els exemples a implementar.

En aquest aspecte, el projecte resultava especialment atractiu, doncs la riquesa final d'aquest vindria donada per la capacitat de traduir els objectius d'un camp d'estudi com la genealogia, mitjançant el grup de dades disponible a través de FamilySearch, en propostes que fossin capaces de satisfer preguntes o inquietuds latents en la societat. A la vegada, aquest mateix aspecte convertia el projecte en aterridor, doncs no seria fins ben entrat en aquest, que comprendríem les opcions disponibles i fins a quin nivell podríem aprofundir en l'elaboració de propostes i exemples.

Un últim aspecte que em va ajudar a decidir-me per aquest projecte va ser la influència de la feina d'interí que estava realitzava en aquell moment com a dissenyador d'experiència d'usuari. El projecte m'oferia la possibilitat d'implementar una pàgina web com a apartat tècnic i aquesta em permetia posar a prova els coneixements d'usabilitat i experiència d'usuari adquirits durant els darrers mesos. De la mateixa forma, s'obria la porta a desenvolupar les habilitats necessàries per programar una pàgina web, una part de la informàtica que no havia explorat durant la carrera i em feia certa gràcia.

## 1.3 Objectius

Tot i la incertesa de quins seran els detalls finals del projecte en el moment de començar, l'abast dels objectius principals sí que es troba ben marcat i definit. Com hem comentat, els objectius principals del projecte giren al voltant de tres grans blocs:

- Estudi de l'API de FamilySearch
  - Estudi de les peces d'informació accessibles i utilitzables.
  - Relacions entre les diferents peces d'informació.
- Bateria d'idees relacionades amb l'API de FamilySearch tenint en compte les restriccions del sistema i la informació disponible per servir com a futurs projectes o com a font d'inspiració.
- Avaluació i implementació d'exemples que es comuniquin o nodreixin de l'API de FamilySearch, n'exposin la informació disponible i ofereixin una idea bàsica de les oportunitats i complicacions que aquesta comporta. Aquest objectiu es divideix en diversos apartats:
  - Estudiar les diferents opcions disponibles per la implementació.
  - Escollir el tipus d'aplicació a desenvolupar i estudiar les tecnologies necessàries per desenvolupar l'aplicació.
  - Definició de l'abast i esquelet de l'aplicació, així com dels exemples que seran desenvolupats i definició dels requisits funcionals i no funcionals de l'aplicació.
  - Implementació de l'aplicació.
  - Procés de certificació i proves del sistema.

Aquest projecte també presenta un seguit d'objectius secundaris o més aviat, objectius personals, que en certa forma m'agradaria deixar plasmats en la memòria.

- Adquisició del coneixement necessari sobre el funcionament d'una pàgina web, quins són els seus components principals i com interactuen.
- Estudi d'algunes de les tecnologies més usades en el mercat actualment.
- Aplicació bàsica dels coneixements d'usabilitat i experiència d'usuari adquirits durant l'etapa d'interí.
- Aprenentatge del llenguatge de maquetació de text LaTeX per tal de formatar articles i documents de caràcter tècnic.

## 1.4 Planificació

L'objectiu d'aquest apartat de la memòria és presentar les diferents planificacions que s'han portat a terme per encarar el projecte en cada una de les convocatòries

matriculades.

### 1.4.1 Planificació Febrer del 2014 - Juliol 2014

Aquest projecte va ser matriculat per primera vegada al Febrer del 2014 amb la intenció de presentar-lo com a principis de juliol del mateix any. La idea inicial era aprofitar el mes de gener per avançar feina i disposar així d'un total de sis mesos per realitzar el projecte.

Al març del 2014 vaig començar a treballar a jornada completa i el projecte va deixar d'avançar a la velocitat esperada. Es van començar a patir forts endarreriments sobre la planificació original fins al punt que el projecte va quedar completament aturat. La figura 1.1 mostra en línies generals la planificació que s'hagués volgut portar a terme en cas de normalitat i ressaltat en vermell la part que es va veure interrompuda.

FASE DEL PROJECTE / MES	Gener 2016	Febrer 2016	Març 2016	Abril 2016	Maig 2016	Juny 2016	Juliol 2016
Estudi de la API							
Generació d'idees per futurs projectes							
Estudi de les tecnologies necessàries							
Implementació de l'aplicació							
Redacció de la memòria							
Preparació de la presentació							

Figure 1.1: Planificació original *Febrer 2014 - Juliol 2014*.

La falta de temps per realitzar un projecte acceptable, conjuntament, a la poca capacitat de maniobra de les que es va disposar entre els mesos de Març i Juliol, va provocar que es descartés la possibilitat de presentar el projecte durant la convocatòria prevista inicialment.

### 1.4.2 Planificació Febrer del 2016 - Setembre 2016

A conseqüència de l'extinció del pla d'enginyeries 2003, el projecte es torna a matricular al Febrer del 2016, tenint en consideració que s'hauria de començar pràcticament de 0.

Donada la diferència de temps entre la primera inscripció i la segona, l'API de FamilySearch s'havia vist sotmesa a grans canvis i la major part del material estudiat i coneixements tècnics adquirits fa dos anys, quedaven completament antiquats.



A pesar de matricular el projecte a mitjans de Febrer es coneixia que aquest no podria ser començat amb agilitat fins a principis d'abril a causa d'una situació excepcional en l'àmbit laboral. Gràcies a la disponibilitat d'una pròrroga extraordinària, que permetia estendre el període d'entrega fins a finals de setembre, la finestra de temps disponible per completar el projecte rondava els cinc o sis mesos.

Cal tenir en compte que la disponibilitat horària en el dia a dia de cara a treballar en el projecte era molt reduïda i en conseqüència, realitzar una bona planificació era essencial si es volien evitar els mateixos problemes que van provocar l'abandonament del projecte en el seu primer intent.

Tant en la figura 1.2, com en les seccions que segueixen a continuació, expliquem com va ser planificada i executada la feina entre els mesos d'abril i setembre.

### **Segona quinzena de Març**

En aquest petit període de temps es va realitzar el primer estudi superficial sobre l'API de FamilySearch amb la finalitat d'observar quins canvis s'havien produït durant els darrers dos anys i com aquests podien afectar o modificar la proposta inicial inscrita del projecte.

L'objectiu d'aquesta repassada ràpida era la de proporcionar una visió global sobre certes limitacions que podrien afectar el desenvolupament del projecte i ens permetés elaborar una planificació coherent de com afrontar i estructurar la feina a realitzar.

### **Primera quinzena d'Abril**

Tot i que l'estudi sobre la informació disponible a través de l'API es trobava en els seus inicis es va aprofitar aquesta quinzena per decidir quina mena d'aplicació volíem implementar. Aquesta decisió obriria pas a la recerca i estudi sobre quines tecnologies serien més adients de cara a la implementació dels exemples i les comunicacions amb l'API de FamilySearch.

També s'aprofitaria aquesta quinzena per familiaritzar-nos amb la diferent documentació disponible sobre l'API de FamilySearch i plantejar-ne l'ordre d'estudi.

### **Segona quinzena d'Abril - Finals de Maig**

Aquest període inicial del projecte resultava crucial de cara a incorporar les eines necessàries al nostre coneixement que ens permetrien completar un dels objectius principals del projecte durant els mesos següents.

Així doncs, l'objectiu era el de detallar i estudiar tots els aspectes referents a la part tècnica de l'aplicació. Escollir de forma correcta era indispensable si volíem evitar tancar-nos portes abans de començar o assegurar-nos de què utilitzàvem eines

26/09/2016									2
19/09/2016									1
12/09/2016								7	
05/09/2016						11.5		6	
29/08/2016					8	11		5	
22/08/2016					7.5	10.5		4	
15/08/2016					7	10		3	
08/08/2016					6.5	9.5		2	
01/08/2016								1	
25/07/2016							9		
18/07/2016				8			8		
11/07/2016				7.5			7		
04/07/2016				7			6		
27/06/2016				6.5			5		
20/06/2016							4		
13/06/2016							3		
06/06/2016					6		2		
30/05/2016					5.5		1		
23/05/2016				6	5				
16/05/2016				5	4.5				
09/05/2016				4	4				
02/05/2016				3	3				
25/04/2016				2	2				
18/04/2016				1	1				
11/04/2016			2						
04/04/2016			1						
28/03/2016	2								
21/03/2016	1								
FASE DEL PROJECTE / SETMANA	Primer estudi superficial de l'API	Decisió tipus d'aplicació a implementar i definició de l'estructura	Decisió i estudi de les tecnologies a utilitzar per l'aplicació web	Propostes de projecte per futurs estudiants i estudi de l'API	Implementació de l'aplicació web	Redacció de la memòria final	Preparació de la defensa del projecte		

Figure 1.2: Planificació final *Febrer 2016 - Setembre 2016*.

eficients que oferien un bon balanç entre esforç i qualitat. Els punts coberts durant aquesta fase van ser:

- Esbrinar els components principals que conformen una pàgina web avui en dia i com interactuen.
- Conèixer les diferents tecnologies disponibles per cada un d'aquests components.
- Escollir del grup de tecnologies estudiat les més adients per fer front als objectius del projecte i estudiar-les a fons.

El segon objectiu d'aquesta fase consistia en estudiar a fons l'API de FamilySearch per obtenir una idea més concreta quina mena de projectes es podrien arribar a realitzar i quins no. Aquest exercici ens ajudaria a comprendre quins exemples tindria més sentit implementar per tal de demostrar la potencialitat i abast de l'API.

### **Mesos de juny i juliol**

Gairebé tot l'esforç durant aquests dos mesos es concentraria en la implementació de l'aplicació web. Es va decidir prioritzar aquesta tasca per sobre de la generació d'idees i l'estudi final de l'API per dos motius.

Gairebé tot l'esforç durant aquests dos mesos es concentraria en la implementació de l'aplicació web. Es va decidir prioritzar aquesta tasca per sobre de la generació d'idees i l'estudi final de l'API per dos motius.

A pesar dels avantatges que oferia desenvolupar l'aplicació en primer lloc, com a contrapartida, també significava tirar endavant una part important del projecte amb el risc de no haver arribat a conèixer la totalitat de l'abast de l'API.

### **Mes d'agost**

L'objectiu del mes d'agost era fer front a tota aquella part del projecte que havia quedat oblidada fins aquest moment. En concret:

- Redactar la memòria del projecte.
- Detallar les diferents propostes de projecte pels futurs estudiants.
- Implementar algunes parts dels continguts estàtics de l'aplicació web.

### **Mes de setembre**

El mes de setembre s'utilitzaria com a marge de maniobra per acabar de tancar aquelles tasques del projecte que poguessin estar sotmeses a petits retards. Segurament, l'acabat de redacció de la memòria i petits retocs en l'aplicació web.

També s'aprofitaria la part final del més, un cop el projecte estigués entregat, per preparar la defensa.

## Secció 2

# La genealogia

Resultaria estrany realitzar un projecte que parla o tracta la genealogia en tots els seus apartats i no realitzar una petita introducció que exposi en què consisteix aquesta ciència.

No representa un objectiu del projecte comprendre l'estat actual de la genealogia en el món contemporani, ni el de crear un dibuix detallat de quines lleis en regulen les seves activitats. No obstant això, sí que es creu que donar una petita visió general dels problemes i preguntes que aquesta ciència pretén abordar pot ajudar a lectors del projecte o futurs estudiants a comprendre millor les limitacions i oportunitats d'aquest sector.

### 2.1 Què és la genealogia?

La Genealogia (del grec: *'genea'*, *'generació'*; i, *'logos'*, *'coneixement'*) és també coneguda pel nom d'història familiar. Aquesta ciència consisteix en l'estudi de les famílies, el seguiment dels seus llinatges, tant ascendents com descendents i l'estudi de la història de les persones.

Els genealogistes, o persones dedicades a la genealogia, tant en l'àmbit privat com personal, utilitzen com a recurs d'investigació arxius històrics rics en dades. Exemples d'aquests recursos poden ser les partides de naixement, documents de defunció, registres d'emigració o altres documents informatius del mateix caire. L'objectiu d'aquests documents és obtenir informació sobre una persona o família per així poder demostrar relacions de parentesc i llinatge o bé, fets empírics relatius a la vida d'un individu en concret.

Un altre recurs que es veu cada cop més utilitzat és l'anàlisi genètic, mètode que té una rebuda, demanda i interès més elevat en l'àmbit personal, que no pas en el científic. La finalitat principal d'aquest mètode és la d'esbrinar relacions familiars passades i presents de l'individu a través de l'anàlisi dels seus gens.

Els motius pels quals una persona pot estar interessada a endinsar-se en el món

de la genealogia són diversos. Un exemple podria ser el desig de situar la seva família en un marc més ampli dins de la història o bé, el sentiment de responsabilitat de cara a preservar la història familiar per les futures generacions.

Els aficionats a la genealogia, que la practiquen com a hobby, generalment investiguen la seva ascendència o la d'una persona propera. Per altra banda, els professionals, acostumen a encarregar-se de realitzar recerques genealògiques per tercers, estudiar i ensenyar mètodes de recerca o mantenir les seves pròpies bases de dades.

Cal entendre que la genealogia no tracta només de recopilar informació sobre el moment històric en què una persona va néixer, viure o morir, sinó també el de recollir informació sobre l'estil de vida que aquella persona va portar, les seves biografies o quins van ser els esdeveniments i motivacions que van conduir i marcar la seva existència. En altres paraules, podríem dir que una part de les preguntes que la genealogia pretén respondre és la de com van viure o quin caràcter van mostrar els nostres avantpassats al viure durant el transcurs d'esdeveniments històrics, com per exemple, la segona guerra mundial.

Voldríem tancar aquesta secció indicant que si l'interès per la genealogia, ha anat en augment en els últims temps, és en gran part gràcies a la digitalització de documents, fet que ha permès que genealogistes amateurs disposin d'un ventall d'eines molt superior al que van disposar els seus avantpassats i per tant, que les possibilitats de mantenir un arbre familiar o realitzar recerca genealògica quedin a l'abast de tothom.

## 2.2 El paper de la genealogia en el transcurs de la història

Com s'ha comentat en l'apartat anterior, avui en dia la genealogia és una ciència que busca en gran mesura respondre preguntes de caràcter personal, no obstant això, aquest no va ser sempre el seu objectiu principal.

Històricament, en les cultures occidentals, les persones estaven interessades a mantenir-se ben informades sobre la seva ascendència de cara a fer latents les seves connexions amb nobles i governants. Generalment, la intenció era protegir la seva situació privilegiada o escapar de la precarietat. En aquesta època, el terme genealogia compartia significat amb el d'heràldica, terme usat avui en dia per la ciència que estudia els escuts d'armes. Així doncs, fins a finals del segle XIX, la genealogia deixava la seva marca en la història com a eina utilitzada principalment per aquells amb drets de poder o riquesa adquirits a través de l'herència.

Aquest exemple, que bé ens podria semblar distant en el temps, no és l'única mostra dels impactes històrics relacionats amb aquesta ciència i com veurem a continuació, existeixen altres exemples molt més propers.

No cal tornar gaires anys enredera per veure com durant l'època de l'alemanya nazi ser capaç de demostrar l'afiliació a la "raça suprema" era necessari per sobreviure o inclòs poder casar-se de forma legal. Per aquest motiu, no ens ha d'estranyar que avui en dia, Alemanya, segueixi sense fer públics la major part dels registres

genealògics del segle XX, doncs els fets històrics han portat a percebre la història familiar com un atac, o amenaça, a la privacitat i seguretat de les persones. Les conseqüències d'aquesta època de la història són conegudes per tothom i un no pot evitar entreveure certes relacions amb el camp de la genealogia.

Per situar un exemple que ens ocupi si pot ser encara més de ple, podem veure el valor de la memòria històrica i dels sentiments d'unió amb els nostres avantpassats arran de la gran quantitat de publicacions i missatges personals, recordant als seus avantpassats i als temps que els va tocar viure, en relació al vuitantè aniversari de l'esclat de la guerra civil espanyola. De fet, Catalunya és un altre clar exemple contemporani, conjuntament amb Alemanya, de com la memòria històrica pot ser present en la cultura, vida i sentiments de bona part d'una nació. Tant en l'àmbit personal, com col·lectiu.

Així doncs, podem concloure que la genealogia, no tant com a ciència sinó com eina, va desenvolupar, desenvolupa i probablement, seguirà desenvolupant, un paper important en la història de la humanitat. No hem d'oblidar que els problemes racials segueixen molt presents en l'actualitat de les nostres societats, Estats Units, n'ha estat últimament un clar exemple, i que és la raça sinó una característica més de les nostres característiques de naixement, o en altres paraules, de les nostres dades genealògiques.

## 2.3 Les lleis reguladores

Les seccions anteriors han introduït i descrit les ocupacions principals de la genealogia en els àmbits professional i amateur, així com el paper d'aquesta en la història. També s'ha esmentat que moltes de les dades amb les quals aquesta ciència interactua són de caràcter personal i per tant, sensibles a un ús impropï si no són regulades i protegides sota certes circumstàncies.

És per aquest motiu que bona part de les dades públiques enregistrades per l'estat, sobretot aquelles que afecten a persones que encara són vives, es troben regulades sota un conjunt de lleis i legislacions. Aquestes lleis varien de nació en nació i per tant, no existeix un estàndard de quina informació és accessible pel domini públic, quina no i sota quines circumstàncies aquesta informació pot ser accedida.

En el cas de l'estat espanyol són dues les principals lleis que regulen l'accés a les dades genealògiques. La llei orgànica de protecció de dades (LOPD) i la legislació consolidada: Llei 20/2011, del 21 de juliol del Registre Civil.

El registre civil espanyol conté informació detallada d'una persona relacionada amb el seu naixement, relacions d'ascendència i descendència, nom i cognoms, emancipació, declaracions de concurs o suspensió de pagaments, nacionalitat, etcètera, etcètera. Com podem veure, aquest registre conté tota mena d'informació sensible i al mateix temps, de gran valor de cara a estudis genealògics.

Els habitants d'Espanya podem demanar accés a l'entrada d'una persona al registre civil mitjançant la presentació d'una sol·licitud digital, escrita o presencial. Per

aconseguir aquesta informació caldrà proporcionar tan dades personals pròpies com el motiu pel qual es vol poder accedir a la partida en concret. Motius recurrents són l'estudi genealògic, gestions administratives o simplement la recaptació d'informació.

Per altra banda, accedir al gruix de la informació no és fàcil i els genealogistes porten xocant amb portes tancades des de fa molts anys. Relacionat amb aquest aspecte, durant l'any 2011 es va aprovar una nova llei del Registre Civil que tenia com a objectiu racionalitzar l'estructura del registre i desjudicialitzar-lo. Aquesta llei havia d'entrar en vigor a partir del 2014, data que va ser posposada fins al juliol del 2015 i recentment ha tornat a ser ajornada fins al 30 de juny del 2017.

La part que farà referència sobre si el nou registre contemplarà l'accés al públic de cara a la recerca genealògica encara està a l'aire, però sembla que hi ha certa esperança gràcies a la inclusió del següent apartat en l'article 80:

4. Amb caràcter excepcional i amb finalitats d'investigació familiar, històrica o científica, es podrà autoritzar l'accés a la informació registral en els termes que reglamentàriament s'estableixin.

Així doncs, sembla que un futur no molt llunyà, aquesta informació podria passar a ser explotable en grans escala de cara a estudis familiars, històrics o científics. Això si, sempre respectant la LOPD.

## 2.4 Els codis ètics en la genealogia

El fet que els genealogistes tinguin accés i treballin amb informació pública, però simultàniament, personal, provoca que la professió es vegi envoltada de codis ètics que tractin de protegir la integritat d'aquesta, el sentiment de professionalitat i la moralitat d'aquells que interactuen amb les dades.

Els codis morals giren al voltant de dos eixos. La protecció de la informació referent a les persones vives i les bones praxis de cara a la manipulació i tractament de les dades.

El primer eix, tal com hem indicat, fa referència a protegir a aquelles persones que encara són vives. En concret, es tracta d'evitar, en la mesura que sigui possible, publicar informació de caràcter personal que pogui resultar compromesa per un individu en concret.

El codi moral que hi ha en el rerefons és que a cap persona li agradaria trobar informació personal publicada al núvol, on tothom la pot accedir, pel simple fet que es tracta d'informació 'pública'. Per tant, cal respectar la privacitat de les persones i no publicar fets o dades compromeses independentment de l'opinió personal del genealogista.

Com ja s'ha mencionat amb anterioritat, la genealogia tracta en gran mesura d'estudiar els nostres avantpassats i la història de tots aquells que van existir abans



que nosaltres, per tant, el xafardeig i la tafaeria no tenen cabuda dins dels codis ètics i morals de la professió.

La segona branca ètica es correspon a un seguit de bones praxis de cara a la utilització d'informació genealògica. Tot i que no es tracta d'un manual oficial, la següent llista de 'regles' serveix per descriure i fer-nos una idea amb un alt grau de fiabilitat, del que significa el concepte de bones praxis en aquesta ciència:

- El genealogista mantindrà les fonts de referència i les citarà quan utilitzi dades fetes públiques per un altre individual o col·lectiu.
- El genealogista no compartirà ni utilitzarà informació no contrastada o amb altes probabilitats de ser errònia.
- El genealogista transmetrà seguretat i confiança a aquells que facin ús dels seus serveis.
- El genealogista donarà suport a aquelles iniciatives que preservin els fitxers públics i l'accés a aquests.
- Es tractarà amb cordialitat i respecte al personal de les facilitats d'investigació.
- El genealogista ajudarà en la mesura que sigui possible als altres genealogistes i organitzacions dedicades a la genealogia.
- El genealogista compartirà els resultats dels seus estudis i investigacions.
- No està permesa la invenció ni exageració de la informació.
- El genealogista complirà amb les lleis en rigor dels conjunts de dades que utilitzarà en els seus estudis.

## 2.5 El procés de recerca genealògica

A tota persona que li interessi realitzar recerca genealògica sobre la seva pròpia família o la d'una persona propera, se li suggereix el següent procés com a mètode de treball.

El procés de recerca es desenvolupa en cicles i cada cicle consta de cinc fases. A continuació es detallen una a una:

1. **Identificar el que es coneix:** Identificar i revisar tota la informació inicial que es coneix. Pel final de la fase s'haurien de tenir recopilats, ordenats i documentats tots els esdeveniments relacionats amb la família o persona a estudiar disponibles.
2. **Decidir que es vol aprendre:** L'objectiu d'aquesta fase és identificar sobre quin individu es vol obtenir informació, que es vol aprendre d'aquesta persona i si és possible el temps i llocs aproximats en els quals aquesta persona va viure.
3. **Seleccionar els arxius a consultar:** Aquesta fase resulta la més complexa de tot el procés. L'objectiu, ordenar de més a menys útils les diferents fonts de dades a consultar i quins arxius resulten més interessants.

4. **Obtenir i consultar els arxius:** Durant aquesta fase es consultaran les fonts de dades seleccionades a l'apartat anterior. Al final d'aquesta, hauríem de tenir anotat tot allò que s'ha descobert i còpies dels documents que suporten els descobriments, ja sigui en format de fotocòpies, notes o qualsevol altra mena de suport físic o digital.
5. **Utilitzar la informació:** Finalment, en aquesta fase, tocarà avaluar la informació descoberta, transportar la nova informació als formularis adients, organitzar la informació i compartir els resultats. Un cop finalitzades totes aquestes tasques s'estarà preparat per tornar a iniciar la roda del procés i seguir així amb la recerca genealògica.

La figura 2.1 mostra les cinc fases d'aquest procés cíclic.

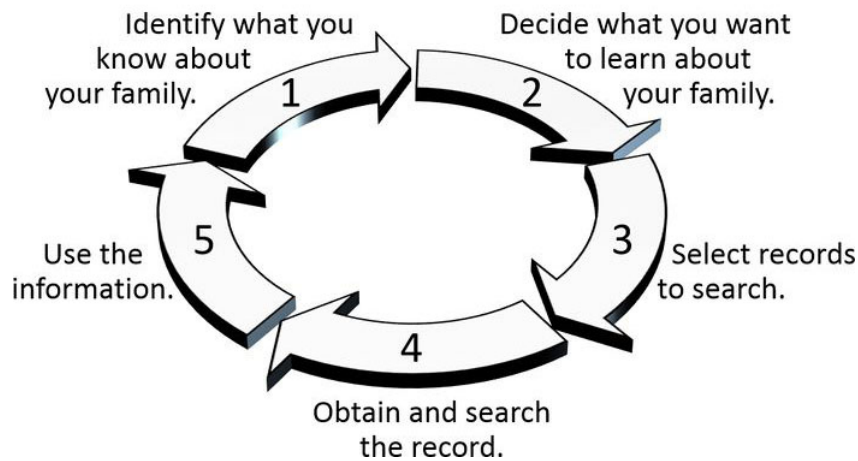


Figure 2.1: Procés de recerca genealògica.

## 2.6 Conclusió

Donem d'aquesta forma per conclosa la breu introducció a la genealogia. Com s'ha pogut observar, es tracta d'una ciència d'especial vocació personal, amb regulacions ambigües i rodejada de codis ètics i morals.

Tot i les dificultats que tant genealogistes amateurs com professionals poden haver de fer front, aquesta ciència és capaç de desemmascarar i ajudar a comprendre les vivències dels nostres avantpassats i crear així un enllaç entre passat, present i futur.

## Secció 3

# L'organització FamilySearch

En aquest apartat de la memòria s'introduirà l'organització de FamilySearch. En concret, s'explicaran els objectius i motivacions sota les que va néixer l'organització, la seva història i el conjunt de funcionalitats i serveis que ofereixen a través del seu portal web i centres d'investigació.

### 3.1 Què és FamilySearch?

Si haguéssim de resumir l'organització en una sola paraula, aquesta seria família. Com s'indicava en la introducció, FamilySearch és una organització sense ànim de lucre destinada connectar famílies a través de diferents generacions. Des de l'organització es creu que les famílies són un condicionador de la felicitat i un dels factors que donen sentit a la vida.

La seva visió com a empresa, amb les seves pròpies paraules, és:

Aprendre dels nostres avantpassats, a través dels llaços familiars, ens ajudarà a comprendre millor qui som, enllaçar el present amb el passat i construir els ponts cap al futur.

Aquesta visió és perseguida mitjançant un equip de professionals i voluntaris que treballen de cara a preservar i compartir el que s'ha convertit en la col·lecció més gran del món d'arxius genealògics. Aquest fet no és fruit de la casualitat, sinó dels més de cent anys que aquesta organització porta recol·lectant, preservant i compartint arxius genealògics des de tots els indrets del globus terraqüi.

FamilySearch organitza i tracta els recursos disponibles tenint present que la finalitat de les dades és poder ser consumides per tercers. Per això, FamilySearch aposta per un model de dades eficaç, on aquelles persones interessades a explorar el seu passat, puguin descobrir amb certa facilitat quin són els seus orígens.

Algunes de les característiques o dades que ajuden a comprendre l'envergadura i èxit d'aquesta organització se citen a continuació:

- Servei gratuït.
- Més de 4 bilions de persones diferents emmagatzemades en el sistema. Per tenir una idea de la magnitud d'aques nombre, actualment s'estima que en el món viuen 7,4 bilions de persones.
- 4.765 centres d'investigació distribuïts arreu del món.
- Servei d'ajuda 24/7.

## 3.2 La història de FamilySearch

FamilySearch, en els seus orígens coneguda com la Societat Genealògica de Utah, va néixer de les mans de l'església de Jesucrist dels Sants dels Darrers Dies l'any 1894. Aquesta església també és coneguda avui en dia pel nom de l'església mormona.

L'organització porta acumulats a les seves espatlles més de cent anys de recerca i preservació d'arxius històrics i genealògics. Durant aquest recorregut, FamilySearch s'ha associat amb més de 10.000 arxius i 200.000 voluntaris de tota mena d'indrets.

Mitjançant l'esforç col·lectiu d'aquestes organitzacions s'ha aconseguit preservar tant índexs com imatges d'arxiu de gran qualitat i posar tots aquests recursos a la disposició de milions de persones de forma gratuïta.

Durant aquest trajecte de més de 100 anys iniciat l'any 1894, l'organització ha tingut l'oportunitat de celebrar grans èxits. Entre ells, destaquen els següents:

- **1938, Microfilm:** La societat genealògica de Utah és pionera en començar a utilitzar microfilm per filmar i emmagatzemar informació relativa a arxius genealògics arreu del món.
- **1942, Family group record archive:** Es crea, de forma manual, una indexació de les genealogies compartides fins aquell moment.
- **1963, Baül d'arxius a les Granite Mountains:** Es completa la creació d'un baül d'arxius amb tecnologia punta situat a les muntanyes pròximes a Salt Lake City, Utah, indret on resideix, avui en dia, la seu de l'organització. Es tracta d'una instal·lació climatitzada que ha estat utilitzada des de la seva creació per preservar còpies de microfilm i arxius digitals de més de cent països diferents.
- **1970, Primers centres d'història familiar:** S'introdueixen els primers centres d'història familiar. Aquests centres formen part d'una ramificació de llibreries que ofereixen accés gratuït a la informació continguda per més de 2,4 milions d'arxius en microfilm. Com hem esmentat en l'apartat anterior, avui en dia existeixen 4.765 d'aquests centres.
- **1985, L'estàndard GEDCOM:** En aquest any FamilySearch introdueix l'estàndard de Comunicació de Dades Genealògiques (GEDCOM). L'estàndard GEDCOM consisteix en un conjunt d'especificacions i regles sobre com s'ha

d'estructurar la informació genealògica de cara a compartir-la amb facilitat a través del núvol.

- **1995, Arbres genealògics digitalitzats:** Es dona l'oportunitat als genealogistes de digitalitzar els arbres genealògics a la seva disposició i habilitar-ne l'accés a altres usuaris.
- **1998, Digitalització d'imatges:** FamilySearch comença a utilitzar tecnologies d'imatge digital per tal de capturar noves fonts de dades i transformar els milions de continguts, emmagatzemats fins ara en microfilm, en imatges digitals. La tecnologia també permet crear índexs de fàcil utilització que relacionen persones amb els continguts digitals.
- **1999, Nova pàgina web:** La pàgina web FamilySearch.org arriba al núvol. En la seva fase inicial, aquesta oferia la possibilitat de cercar informació en els registres històrics de forma relativament simple.
- **2012, Noves tecnologies digitals:** S'incorpora a les tecnologies utilitzades per FamilySearch la tecnologia dCamX, utilitzada per la digitalització de documents i la creació de sales de lectura digital, responsables de facilitar les tasques de comunicació i alliberació de coneixement entre els diferents centres.

Un dels altres grans èxits de l'organització, del que malauradament no es coneix la data exacte de creació, va ser **l'obertura de la FamilySearch API**.

Aquesta va suposar que aplicacions externes poguessin connectar-se a les bases de dades de FamilySearch i utilitzar-ne la informació d'una forma regulada i eficient. Resulta prou evident que sense l'existència d'aquesta API, aquest projecte mai hagués pogut tenir lloc.

### 3.3 L'església mormona i la família

Ja que el principal benefactor de FamilySearch és l'església mormona, creiem que és interessant estudiar de forma breu els orígens d'aquesta.

L'església de Jesucrist dels Sants dels Darrers Dies és una església que considera que la religió hauria de tornar als seus inicis apostòlics. Va ser fundada pel nord-americà Joseph Smith el 6 d'abril del 1830, a l'oest de Nova York.

Considerada actualment com la quarta comunitat cristiana més gran als Estats Units, troba situada la seva seu en l'actualitat a Salt Lake City, Utah. No obstant això, durant els inicis de l'església, Smith tenia la intenció de crear la Nova Jerusalem a prop de Nova York, en una ciutat que anomenaria *Zion*.

L'església, amb origen als voltants de Nova York, es va desplaçar cap a Kirtland, Ohio, des d'on va començar a expandir-se per Jackson County, Missouri, terra en què Smith volia situar la seu del col·lectiu en un futur pròxim.

Joseph va veure contrariats els seus plans, quan l'any 1833, els colons van expulsar brutalment al col·lectiu de Missouri. Com que no disposaven dels recursos

militars necessaris per recuperar el territori per la força, es van veure obligats a anar desplaçant-se al llarg de diferents localitzacions, sempre per culpa de conflictes amb els nadius de les terres, fins a establir-se a Nauvoo, Illinois.

Després de la mort de Smith, per tal d'evitar els conflictes armats amb els residents d'Illinois, el col·lectiu es va desplaçar cap a Nebraska i més endavant, durant l'any 1847, a les terres que serien conegudes com a Utah.

Durant aquesta època, l'església es va veure sotmesa a grans pressions i crítiques a causa de la seva tolerància per la poligàmia. Les tensions entre el col·lectiu i el govern d'Estats Units anirien en augment, fins que l'any 1890, el congrés va disgregar l'església i es va apoderar de molts dels seus béns.

Arribats aquest punt, l'església fundada per Smith, va decidir deixar de donar suport als matrimonis plurals, però sense desfer les famílies que ja es trobaven unides sota aquestes condicions.

Durant el segle XX, l'església va créixer substancialment i es va veure sotmesa a un procés d'internacionalització, en gran mesura, gràcies a la feina dels missioners enviats a diferents indrets del món.

Durant aquest període el col·lectiu es va convertir en un ferm defensor de les famílies nuclears, és a dir, de les famílies que consisteixen en dos progenitors i la seva descendència. L'església també va oposar-se en aquesta època a l'esmena pels drets igualitaris entre homes i dones, els casaments entre persones del mateix sexe i l'eutanàsia.

Hem volgut redactar aquests paràgrafs previs sobre els orígens de l'església mormònica per tal de poder presentar, amb cert rigor històric, com l'església va veure canviat i evolucionat el concepte de família al llarg del temps.

També queda latent, d'aquesta forma, com la història del col·lectiu es va veure marcada pel rebuig i el desterrament de moltes terres, fins al punt que van haver de recórrer a l'ús de missioners, repartits arreu del món, per tal de sobreviure com a religió.

Així doncs, creiem que per aquest projecte no esdevé necessari entrar en més detall pel que fa a les doctrines i pràctiques de l'església, ni enumerar quines són les principals diferències entre l'església mormona i les altres corrents del cristianisme. Per altra banda, sí que volem realitzar una reflexió final sobre la posició actual de l'església mormona respecte a la família.

Pels mormons, les famílies representen els lligams que uneixen a les persones en relacions personals i les connecten tant amb les passades com amb les futures generacions. Creuen que cap èxit en la vida, pot compensar el fracàs en l'àmbit familiar.

Segons el seu punt de vista, construir nuclis familiars units i forts és el remei a molts dels fracassos que tenim les persones com a societat i creuen que la família inspira a l'individu a pensar més enllà de l'interès propi o la gratificació immediata i l'ànima a entregar-se per altres persones, comunitats i a déu.

Forma part també de la cultura mormona la pràctica o deure d'acumular i preservar, tant les històries dels seus avantpassats com les pròpies, en benefici d'aquells que encara estan per arribar, enllaçant, d'aquesta forma, generacions desconnectades d'una altra forma.

Entenen la naturalesa real de la família, com un algú que transcendeix l'aquí i l'ara i que permet a les persones extreure forces d'aquells que van viure abans que nosaltres.

Concloïen, si ajuntem les dues variables que van marcar l'esdevenir de l'església mormona fins als temps contemporanis, és a dir, la seva semi forçada internacionalització i la importància del nucli familiar en la seva cultura, no hauríem de mostrar-nos sorpresos pel fet que el col·lectiu s'hagi convertit en un dels referents mundials en el camp de la genealogia.

## 3.4 Serveis per organitzacions amb arxius genealògics

Com ja s'ha comentat en seccions anteriors, FamilySearch no és només una organització dedicada a posar a disposició del públic registres genealògics, sinó que també pretenen ajudar a altres organitzacions a digitalitzar i publicar els seus documents al núvol de forma econòmica, ja sigui mitjançant la plataforma FamilySearch o la construcció de noves eines pròpies al núvol.

Sigui com sigui, FamilySearch ofereix cinc serveis a altres organitzacions genealògiques:

### 3.4.1 Captura d'imatges

Obtenir imatges de qualitat és normalment el procés més costós per aquelles organitzacions que volen digitalitzar els seus registres, tant en el sentit econòmic, com en base als recursos humans necessaris.

El microfilm, que fins fa poc era l'estàndard en la indústria, comença a cedir pas al món digital i tant si l'objectiu de les organitzacions és digitalitzar el seu contingut mitjançant medis propis o utilitzant la tecnologia disponible en els centres de recerca de FamilySearch, aquests ofereixen la seva ajuda a les organitzacions que la sol·licitin.

La tecnologia dCamX, utilitzada per FamilySearch, es caracteritza per la creació d'imatges d'alta qualitat, de forma eficaç, al mateix temps que es capturen metadades de la imatge. Aquest aspecte, conjuntament amb un procés de publicació posterior fàcil i ràpid, fan que aquesta tecnologia estigui cridada a ser el nou estàndard a la indústria.

### 3.4.2 Conversió de formats digital

Aquest servei està pensat per aquelles empreses o organitzacions que ja disposen d'una elevada quantitat de material en format de microfilm.

La tecnologia digital ha canviat dràsticament com els registres són capturats, emmagatzemats i fets accessibles. Aquesta tecnologia segueix progressant i per tant resulta indispensable començar a adaptar-se al més aviat possible.

FamilySearch posa a disposició de les organitzacions genealògiques la possibilitat d'utilitzar els mateixos processos i software que utilitzen ells per digitalitzar la seva col·lecció de més de 2,4 milions de microfilms. FamilySearch, també ofereix la possibilitat d'emmagatzemar els fitxers digitals d'aquestes organitzacions en els seus servidors, un cop convertits, si així ho prefereixen.

### 3.4.3 Indexació en línea

Un cop un registre ha estat digitalitzat en forma d'imatge, la informació principal necessita ser extreta i transcrita per tal de poder produir índexs sobre els quals clients o usuaris puguin cercar.

L'aplicació d'indexació en línea, creada per FamilySearch, permet, mitjançant la cadena de voluntaris, crear índexs de forma ràpida i precisa. Els arxius de les organitzacions, que així ho sol·licitin, podran disposar d'accés a aquesta cadena de voluntaris per digitalitzar els seus índexs o accés a les eines d'indexació auxiliars que permeten la creació de projectes propis.

### 3.4.4 Accés en línea

Si un document o registre no esdevé fàcilment accessible, resulta de poc valor pels usuaris. FamilySearch ofereix dos serveis diferents depenent de si les organitzacions desitgen fer públic l'accés a les seves dades a través de FamilySearch.org o no.

En cas de voler per públics els registres, FamilySearch s'ofereix a penjar i mantenir els registres de forma econòmica. En cas de voler mantenir els registres en un àmbit privat, l'organització posa a disposició dels interessats les eines i experiència necessàries per crear un espai propi al núvol.

### 3.4.5 Preservació dels registres i fitxers físics

FamilySearch ofereix l'opció a les organitzacions de custodiar còpies de seguretat dels seus fitxers, en el baül de tecnologia punta situat a las Granite Mountains. En l'actualitat, còpies d'arxius de microfilm i digitals, provinents de més de cent països diferents, es troben guardades en aquest baül per precaució.

El fet de disposar de còpies de seguretat pels fitxers genealògics, suposa la salvació de registres en cas de terratrèmols, incendis, inundacions, tornados, guerres i actes



humans no controlables en les seues oficials dels arxius.

Les mesures de seguretat que FamilySearch utilitza pels seus registres i que a la vegada, queden a disposició d'altres organitzacions, són:

- Processos complets i automatitzats de comprovació, validació i actualització dels registres per garantir la màxima protecció possible.
- Migració gradual i eficient de registres cap a noves tecnologies quan els formats previs quedin obsolets, garantint així la seva accessibilitat a llarg termini.
- Processos de conversió i preservació de dades que compleixen amb les regulacions sobre els Sistemes de Registres d'Informació Oberts (OAIS).
- Col·leccions d'informació emmagatzemada al núvol i distribuïdes en diferents clústers arreu del món per garantir una alta capacitat d'emmagatzematge, escalabilitat i protecció contra els desastres.
- Utilització de les últimes tecnologies en el tractament de dades d'alta densitat.
- Procés ràpid i eficient de cara a processar l'arribada de nous registres al sistema. El sistema actual és capaç de processar més de vint terabytes al dia, mesura que incrementa, al mateix temps que la tecnologia avança.
- Servidors configurats en clústers virtuals per garantir una escalabilitat infinita.
- Facilitat amb control climàtic, prevenció de focs, fonts d'energia auxiliars per casos d'emergència i replicació d'arxius digitals.

#### 3.4.6 Conclusions sobre els serveis professionals

En els apartats anteriors s'ha pogut observar com FamilySearch està clarament interessada a posar les seues tecnologies a disposició d'altres organitzacions genealògiques.

Aquesta estratègia de col·laboració els permet incorporar a les seues bases de dades registres d'informació, d'altre forma inaccessibles i garantir la persistència de les dades davant d'esdeveniments no controlables, d'informació gestionada per altres organitzacions.

### 3.5 Serveis per particulars i aficionats a la genealogia

A part dels serveis per organitzacions, FamilySearch disposa d'un ampli ventall d'eines i funcionalitats pels genealogistes amateurs o aficionats. Aquestes eines es poden dividir en dos grans blocs. El bloc encarregat de gestionar la informació genealògica dels usuaris i el bloc de cerca sobre les bases de dades de FamilySearch.

Les eines de gestió, sobre la informació genealògica, permeten als usuaris digitalitzar la seua informació d'una forma fàcil, estructurada i accessible que compleix amb els estàndards del sector.

Per altra banda, les eines de cerca, constitueixen el principal atractiu de l'organització per particulars i aficionats a la genealogia. Aquest conjunt d'eines han estat pensades per l'exploració de les dades emmagatzemades en l'ampli catàleg de FamilySearch.

Finalment, FamilySearch també permet, a aquelles persones interessades, formar part d'un encadenat de voluntaris lligat a l'organització. Aquest grup de voluntaris s'encarrega d'ajudar, coordinar i arbitrar la indexació de registres genealògics.

### 3.5.1 Eines genealògiques

El primer bloc d'eines al que fèiem referència en la introducció, eren el conjunt d'eines encarregades de la manipulació de la informació genealògica. Les eines principals giren al voltant de la creació d'arbres genealògics, amb opcions per adjuntar fonts d'informació, arxius multimèdia, documents i altres tipus de documents.

Actualment, existeixen moltes eines que satisfan aquesta mateixa necessitat, tant en línia, com aplicacions per escriptori i mòbil; tanmateix, no són moltes les que ofereixen la possibilitat de completar la informació coneguda pels usuaris amb registres i col·leccions penjades al núvol.

Dins d'aquest grup més reduït, destaquen Ancestry.com i FamilySearch.org, on només la darrera és d'accés gratuït. El conjunt de funcionalitats, que destaquen per part de FamilySearch, es detallen breument a continuació.

#### Arbre familiar

Dins de l'estudi de la genealogia, els arbres genealògics o arbres familiars solen ser l'element més conegut o visual per representar els lligams familiars i l'història familiar d'un individu.

Els arbres familiars a FamilySearch són creats mitjançant les relacions de parentesc entre diferents persones. Aquest fet és important, ja que FamilySearch demana crear aquestes persones per tal d'incorporar-les a les seves bases de dades i fer-les públiques a menys que s'especifiqui que volen ser utilitzades només privadament.

La peculiaritat de FamilySearch és que quan es crea una nova persona, es busca en el sistema, persones que podrien encaixar amb la persona proporcionada. D'aquesta forma, es redueix la creació de persones duplicades i inclús, a vegades, la informació proporcionada és complementada amb la ja existent en l'arbre familiar de FamilySearch.

Un exemple d'arbre familiar de FamilySearch pot ser visualitzat en la figura 3.1.

#### Fan chart

El Fan chart, o gràfic en forma de ventall, ofereix la possibilitat de visualitzar l'arbre familiar d'una persona, de forma més compacta. Aquest format ofereix una visió de 360 graus només sobre l'ascendència de la persona i de la seva parella.

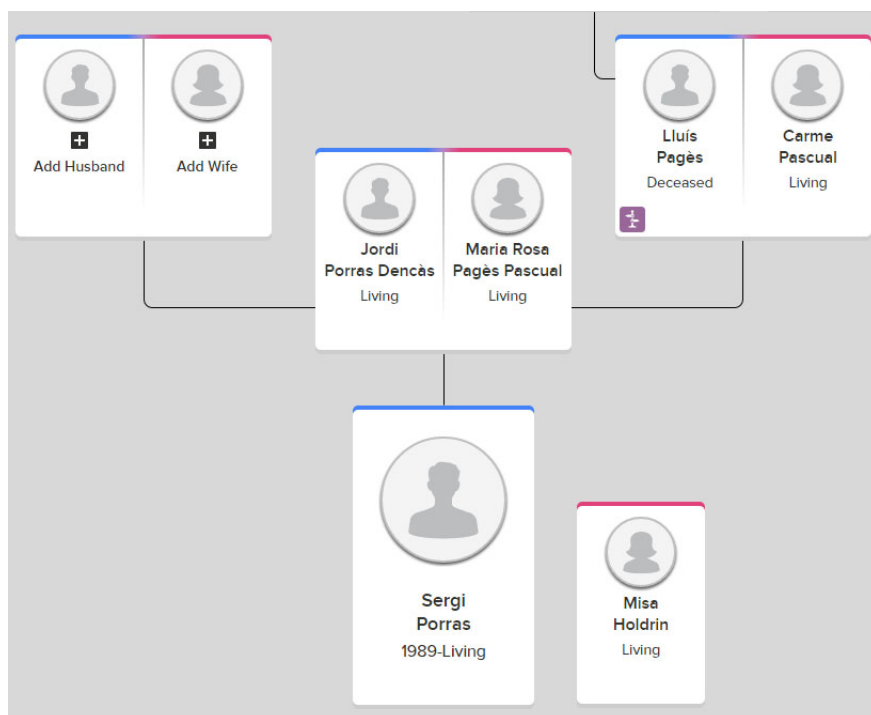


Figure 3.1: Exemple d'arbre familiar.

Oferim un exemple d'aquesta visualització en la figura 3.2.

### Detalls personals

Una altra característica de les eines genealògiques és la capacitat de complementar la informació genealògica d'un usuari amb informació personal o informació genealògica més detallada.

El conjunt d'informació extra que pot ser introduïda, s'exposa a continuació:

- **Esbós de la vida d'una persona:** Resum, d'estructura narrativa, que detalla com va viure la persona i quines van ser les diferents etapes de la seva vida.
- **Informació vital:** Gran part d'aquesta informació és la que ja s'ha introduït en el moment de crear la persona a l'arbre genealògic. Aquesta secció permet editar-la i afegir esdeveniments importants en la vida d'una persona, més enllà de la informació bàsica de naixement o defunció.
- **Fonts d'informació:** Permet introduir informació relacionada a les fonts d'informació que contrasten i validen la informació introduïda sobre la persona.
- **Discussions:** Les discussions són conversacions que poden obrir els usuaris per tal de discutir informació relativa a la persona.



Figure 3.2: Exemple de Fan Chart.

- **Notes:** Les notes permeten anotar detalls concrets o informació extra que no té sentit dins de cap altre apartat.

## Memòries

Les memòries són en gran part contingut multimèdia que els usuaris poden penjar i relacionar amb les persones de l'arbre familiar per tal d'aportar informació personal extra sobre aquestes.

Si recordem la definició de genealogia, aquesta no tractava només sobre l'estudi dels llinatges familiars, sinó també sobre l'estudi de com van viure aquestes persones. Aquest conjunt d'arxius multimèdia, permeten complementar la informació genealògica coneguda amb informació que ajudi a comprendre qui van ser aquestes persones o que va caracteritzar la seva vida.

FamilySearch suporta quatre tipus d'artefactes diferents. A continuació s'explica en més detall en que consisteix cada una d'elles:

- **Fotografies:** Les fotografies representen la ciència o art encarregat de crear imatges perdurables en el temps de persones, moments o esdeveniments. Aquests artefactes permeten adjuntar imatges d'una persona en el seu arbre familiar.
- **Documents:** Els documents solen ser fitxers PDF que proporcionen informació extra sobre aspectes concrets de la persona. Per exemple, un testament.

- **Notes de veu:** Les notes de veu són arxius d'àudio que els usuaris relacionin amb la vida d'una persona en qüestió. Un exemple podria ser, per exemple, un missatge de contestador o gravació auditiva.
- **Històries:** Les històries són peces narratives que els usuaris poden escriure per relatar les vivències d'una persona o documents escrits per la mateixa persona representada a l'arbre familiar.

### Family Booklet

El Family Booklet és un document que FamilySearch crea sota petició que recopila tota la informació disponible relacionada amb la família de la persona que el sol·licita. Aquesta informació s'utilitza per crear un petit llibret que pot ser encarregat en format físic o descarregat digitalment.

## 3.6 Recerca genealògica

El segon bloc d'eines que FamilySearch posa a disposició dels usuaris i aficionats a la genealogia són les eines necessàries per cercar en el seu immens catàleg de registres, col·leccions, genealogies i llibres genealògics.

Existeixen diferents funcionalitats de cerca segons el tipus d'informació que es vol cercar. A continuació detallem la informació principal de cada una.

### 3.6.1 Cerca de registres

La cerca de registres és sens dubte la principal funcionalitat de cerca que queda a disposició dels usuaris. Aquesta permet cercar avantpassats ja difunts en els registres històrics de FamilySearch.

La cerca pot ser delimitada mitjançant diferents paràmetres de cerca i refinada posteriorment mitjançant l'ús de filtres. Els principals paràmetres amb els quals es pot configurar la cerca són:

- Nom i cognoms
- Cercar pel lloc i data aproximada o exacte d'esdeveniments particulars.
- Nom i cognoms dels relatius més propers com poden ser per exemple els pares o la parella.
- Restringir per localització del registre o persona. Pot ser una restricció de continent, país, província, ciutat, etcètera.
- Restringir per tipus de registre. Per exemple, obtenir només registres de naixement, baptisme, casament, servei militar, etcètera.
- Restringir per número de lot.

- Restringir per número de microfilm.
- Nivell d'exactitud desitjat sobre els camps introduïts. Es pot decidir entre un nivell més lax o un que intenti satisfer totes les condicions que han estat introduïdes.
- Restricció per col·lecció. Una col·lecció és una font de dades en concret. Per exemple, el cens de Nova York del 1905.

La figura 3.3 mostre el formulari de cerca bàsic.

**Search Historical Records**

Search for a deceased ancestor in historical records to uncover vital information from their life.

**DECEASED ANCESTOR'S NAME**

First Names  Last Names

**SEARCH WITH A LIFE EVENT:**

Birth | Marriage | Residence | Death | Any

Birthplace  Birth Year (Range) From  To

**SEARCH WITH A RELATIONSHIP:**

Spouse | Parents | Other Person

Spouse's First Names  Spouse's Last Names

**RESTRICT RECORDS BY:**

Location | Type | Batch Number | Film Number

Country  State or Province

☐ Match all terms exactly

Figure 3.3: Exemple del cercador de registres.

### 3.6.2 Cerca de genealogies

La cerca de genealogies consisteix en la cerca sobre els arbres de família pujats i creats a FamilySearch per altres usuaris o provinents de registres oficials. La fiabilitat de les línies familiars creades per usuaris, varia d'arbre en arbre i els usuaris estan més que convidats a comprovar la veracitat de les genealogies resultants de la cerca.

La cerca de genealogies es pot concretar mitjançant els següents paràmetres:

- Nom i cognoms
- Cercar pel lloc i data aproximada o exacte d'esdeveniments particulars.

- Nom i cognoms dels relatius més propers com poden ser els pares o la parella.
- Restringir els arbres genealògics retornats a aquells que han estat creats per usuaris o importats de diferents sistemes i arxius oficials.

### 3.6.3 Cerca per catàleg

Els catàlegs són materials genealògics com poden ser llibres, materials en línia, pel·lícules de microfilm, microfiche i altres publicacions, com per exemple, revistes. Molts d'aquests recursos poden ser agafats en préstec en els centres d'història de FamilySearch.

### 3.6.4 Cerca en llibres d'història genealògica

La col·lecció de llibres d'història genealògica consisteix en més de 200.000 publicacions digitalitzades provinents de les més importants llibreries d'història familiar existents. La col·lecció inclou, evidentment, històries de família, revistes genealògiques, guies d'iniciació a la recerca genealògica, diccionaris geogràfics, històries medievals i arbres genealògics.

### 3.6.5 Wiki de FamilySearch

La wiki de FamilySearch és una petita enciclopèdia que pretén assistir, sobretot als nou vinguts en el món de la recerca genealògica, a comprendre com conduir i enfocar la recerca depenent de les preguntes o informació que estiguin intentant respondre o trobar.

De la wiki cal destacar la informació disponible per cada regió, país o província. Per cada un d'aquests nivells, es disposa d'informació sobre quins poden ser els punts o registres d'entrada més interessants segons la informació que s'estigui cercant.

## 3.7 Projectes d'indexació

La secció d'indexació posa a disposició dels usuaris tota la informació necessària per convertir-se en voluntaris i començar a transcriure informació continguda en registres genealògics digitalitzats.

El procés d'indexació es realitza a través d'un programa creat per FamilySearch que pot ser descarregat des de la mateixa web i s'encarrega de gestionar els registres que l'usuari pot indexar segons els projectes en els quals aquest es trobi inscrit.

Al mateix temps, FamilySearch està a punt de treure un nou servei que permetrà la indexació en línia des del mateix navegador.

### Procés d'indexació

El cicle de vida d'un registre consisteix en les següents fases:

1. Digitalització del registre: En aquesta fase els arxius i registres es digitalitzen.
2. Agrupació dels registres en grups de 20–50 camps diferents dels que cal extreure informació.
3. Dos voluntaris diferents transcriuen els valors dels camps digitalitzats.
4. Si la informació introduïda pels dos voluntaris no coincideix, un àrbitre avalua les dues entrades i pren una decisió.
5. Si les dues extraccions coincideixen, o un cop han estat validades per un àrbitre, les dades s'envien a una base de dades i són preparades per la seva publicació.

Transcriure dades, sobretot aquelles que provenen de documents antics, no és simple i en conseqüència, el programa d'indexació permet introduir com a resposta que part dels camps del registre no poden ser transcrits o que no s'entén el que posa.

De la mateixa forma, FamilySearch proporciona exemples d'escriptura antiga, en diferents idiomes, per ajudar als usuaris a desxifrar els camps més complicats.

### Què pot indexar un voluntari?

Els voluntaris poden escollir un o varis dels diferents projectes que actualment es troben en el procés d'indexació i participar en ells. L'usuari pot escollir amb completa llibertat si vol treballar en projectes d'un país o llengua específica i la quantitat de temps que vol dedicar-hi.

## 3.8 Conclusió sobre les eines per particulars

Tot sembla indicar que FamilySearch està enfocant una part del seu portal web en oferir eines atractives pels usuaris amb l'objectiu de capturar informació genealògica, d'altre forma, difícilment accessible.

Al mateix temps, l'organització està explorant mètodes alternatius i diferents per tal d'aconseguir ampliar la seva col·lecció de registres digitals de la forma més ràpida i eficaç possible. Per aconseguir-ho, està implicant als seus usuaris en el procés d'indexació, principal coll d'ampolla en la digitalització de documents.

Volem recordar, en aquesta conclusió, que molts cops la informació genealògica no passa a estar disponible al domini públic fins molts anys després de la defunció dels seus individus i en molts altres casos, mai arriba a veure la llum.

És per això, que capturar aquesta informació de primera mà, a través dels usuaris, redueix l'espera necessària per aconseguir les dades i redueix les possibilitats d'informació incorrecta sobre aquestes.



## Secció 4

# Introducció a l'API de FamilySearch

### 4.1 El portal de desenvolupadors

Tota la informació disponible per tal de poder començar a familiaritzar-se amb l'API de FamilySearch, pot ser trobada en el portal de desenvolupadors.

Aquest apartat de la web està format per diferents seccions, malauradament, l'estructura no acaba de resultar del tot clara per una persona que vulgui iniciar-se per primer cop en l'ús d'aquesta API.

Si ens enfoquem més en la documentació disponible, que no pas en l'estructura proposada per l'organització, podem veure que la informació es podria distribuir, en certa forma, en els següents grups:

- **Requisits tècnics:** Conjunt d'informació necessària per comprendre l'estructura de l'API, els formats de dades que maneja i els passos necessaris per començar a interactuar amb aquesta.
- **Recursos disponibles i rutes d'accés:** Informació detallada sobre cada recurs accessible a través de l'API. En concret, disposa dels detalls de com accedir al recurs, les operacions que es poden realitzar sobre ell, la informació que conté i quines són les connexions amb altres recursos.
- **Evolució i canvis produïts a l'API:** Informació semi ordenada de com l'API s'ha vist evolucionada al llarg del temps i un recull dels canvis produïts sobre els recursos, procés de certificació, material de documentació i eines de desenvolupament.
- **Serveis extres oferts per l'API:** Aquest recull d'articles conceptualitza característiques de l'API com poden ser els recursos d'*emmagatzematge*, *localització* o *throttling*.
- **Eines de desenvolupament:** Recull d'entorns de desenvolupament i eines

extres que poden facilitar la feina del desenvolupador.

- **Certificació:** Recull la informació necessària per gestionar els diferents processos de certificació i informació sobre les regulacions a les quals s'ha de fer front en cas de voler certificar l'aplicació.

## 4.2 L'arquitectura de l'API

### 4.2.1 Què és una API?

Abans d'entrar en detall en com funciona una API, estaria bé definir, amb una mica més de precisió, en què consisteix exactament.

Una API, de l'anglès 'Application Programming Interface' o Interfície de programació d'aplicacions en català, representa el conjunt de subrutines, funcions i procediments que ofereix una biblioteca per tal de ser utilitzada en el software de tercers com una capa d'abstracció. Aquest conjunt de subrutines, funcions i procediments, acostumen a oferir accés a certs serveis o conjunts de dades d'un particular, a tercers, de forma controlada.

### 4.2.2 L'arquitectura REST

L'arquitectura sobre la qual està creada l'API de FamilySearch és una arquitectura REST. Les sigles provenen de l'anglès i representen el concepte: Representational State Transfer (REST).

Les architectures REST es caracteritzen per estar orientades els recursos més que a les accions que es poden realitzar sobre ells i com a peculiaritat, es caracteritzen per sis regles o restriccions.

Aquestes són: interfície uniforme, sense estat, client-servidor, emmagatzemables, sistema per capes, i codi sota petició. Aquests sis conceptes són doncs els que defineixen les bases de es architectures REST.

És diu que una arquitectura REST és orientada als recursos perquè estan construïdes al voltant d'objectes i les relacions entre aquests, en comptes d'accions. Per exemple, a l'API de FamilySearch, es parla de persones i esdeveniments, en comptes de llegir persones o crear esdeveniments i en canvi, aquestes operacions, passen a formar part dels objectes *Persona* i *Esdeveniment*.

L'intercanvi de dades es produeix mitjançant l'ús de diferents representacions. Aquestes, expliquen com els recursos són tractats per l'API i de quina forma han de ser realitzades les comunicacions entre el servidor i el client. Els formats més freqüents són JSON i XML. FamilySearch ofereix suport per ambdós formats.

Per posar un exemple reduït que il·lustri el que estem explicant, podem descriure de la següent forma, el que podria ser una operació contra l'API de FamilySearch,

especificant quin element seria considerat el Recurs, quin el Servei i quin la Representació:

- **Recurs:** Persona (informació relacionada amb una persona en concret)
- **Servei:** Obtenir informació de la persona (GET)
- **Representació:** Nom, cognoms, esdeveniments relacionats amb la vida de la persona, etcètera, en format Llengatge de Marcatge Extensible (XML) o Notació d'Objectes Javascript (JSON).

Com també s'ha comentat, l'arquitectura REST es caracteritza per la implementació de sis restriccions imposades sobre el sistema. A continuació s'exposa amb més detall, cada una d'elles.

### Interfície uniforme (uniform interface)

Aquesta restricció s'encarrega de definir la interfície de comunicació entre el client i el servidor.

En una arquitectura REST, s'utilitzen els protocols de comunicació HTTP i HTTPS de forma conjunta amb els Identificadors de Recurs Uniforme (URI), per aconseguir accés als diferents recursos i operacions proporcionades per l'API.

Els verbs permesos pels protocols de comunicació web són els coneguts: get, put, post, delete, options and head.

Per exemple, per fer la petició de lectura sobre el recurs d'una Persona a l'API de FamilySearch, executariem la següent crida HTTP o HTTPS mitjançant el verb i URI especificats a continuació:

GET /platform/genealogies/persons/2:2:PPPJ-MYZ7.

### Sense estat (stateless)

Aquesta restricció implica que el servidor no emmagatzema la informació del client. Això implica que cada petició d'aquest cap a l'API, ha de contenir tota la informació necessària perquè el servidor l'identifiqui i es defineixin les regles de comunicació adequades per processar la petició.

Hi ha exemples d'operacions a l'API de FamilySearch, com per exemple el procés d'identificació OAuth V2, que no són realment RESTful, doncs aquestes sí que guarden informació del client durant les diferents parts de la comunicació.

### Client-servidor (client-server)

Per comprendre aquesta restricció, cal comprendre primer en què consisteix un sistema desconnectat.

En el cas de les arquitectures REST, un sistema desconnectat implica que el client mai tindrà accés directe a les bases de dades que emmagatzemen la informació, i que per tant, sempre haurà d'accedir a les dades mitjançant l'intermediari. En aquest cas, l'API.

Els protocols de comunicació descrits prèviament (HTTP i HTTPS), i la interfície de comunicació, són els encarregats de gestionar les comunicacions entre client i servidor.

### Emmagatzematge en el client (cacheable)

Aquesta restricció fa referència a si les respostes retornades, des del servidor, al client, poden ser emmagatzemades per aquest i durant quant de temps les pot guardar. Existeixen tres nivells de configuració diferents:

- **Implícit:** Si és el client el que decideix quant de temps guardarà les dades o informació retornada, es tracte d'emmagatzematge implícit.
- **Explícit:** Si és el servidor el que mana i posa les regles, parlem d'emmagatzematge explícit.
- **Negociat:** Quan el client i el servidor negocien i arriben a un acord, es tracta d'emmagatzematge negociat.

### Sistema per capes (layered system)

Aquest principi, o restricció, es basa en el fet que el client no pot assumir que tindrà connexió directa amb el servidor. És a dir, poden existir diferents intermediaris en forma de hardware i software entre client i servidor.

Això, facilita l'escalabilitat i persistència del sistema gràcies al fet que el client no s'ha de preocupar de comunicar-se amb elements o tecnologies específiques. D'aquesta forma, el servidor es pot veure subjectes a canvis de forma transparent pels clients.

### Codi sota petició (code on demand)

Restricció que regula com, de forma excepcional, el servidor pot proporcionar accés al client sobre certes parts de la lògica del funcionament. Alguns exemples poden ser els *Java Applets* o blocs de codi *JavaScript*.

## 4.3 Formats de dades utilitzats pel Sistema

FamilySearch utilitza tres formats de dades diferents per representar la informació emmagatzemada en les seves bases de dades i dos formats extres per codificar aquesta informació i enviar-la a través del núvol.

Els conjunts de dades utilitzats per representar els recursos són els que segueixen:

- Les dades genealògiques es representen mitjançant el format GEDCOM X.
- Els recursos o objectes específics del model de FamilySearch, es representen mitjançant una extensió del model de dades GEDCOM X.
- El format de dades Atom, o atòmic, s'utilitza per proporcionar un format simple per les meta-dades.

#### 4.3.1 El format de dades GEDCOM i GEDCOM X

El terme GEDCOM, és un acrònim de l'anglès Genealogical Data Communications.

El format GEDCOM consisteix en un conjunt de regles d'aplicació per tal de representar informació genealògica. Aquest format de dades, creat per FamilySearch l'any 1984, s'ha convertit en l'estàndard de la indústria.

Per simplificar-ho, podríem entendre un fitxer en format GEDCOM, com un fitxer de text que emmagatzema informació genealògica d'una persona i les metadades necessàries per poder enllaçar als diferents fitxers de la mateixa persona.

Tot i que l'última versió, datada del 1996, segueix sent molt utilitzada, FamilySearch va proposar durant l'any 2012, canviar aquest estàndard per la seva nova versió anomenada GEDCOM X.

El format de dades per la Comunicació de Dades Genealògiques Extesa (GEDCOM X), representava un nou projecte de codi obert i es diferenciava del seu antecessor en la implementació d'un sistema que facilitava la inclusió d'arbres genealògics i fonts de dades als recursos ja existents.

Al mateix temps, el nou estàndard també donava suport a l'intercanvi i enllaçament de dades a través del núvol i es creava així la primera versió de l'API de FamilySearch.

En la taula 4.1, s'ofereix un petit exemple de com la part bàsica del recurs persona és codificat sota el format de dades GEDCOM X.

Table 4.1: Codificació GEDCOM X del recurs Persona

Nom	Descripció	Format de dades	Restriccions
Private	Indica si la instància de la persona ha estat designada com a privada o pública.	Booleà	OPCIONAL. Una descripció de com les aplicacions han de tractar les dades de caràcter privat.
Gender	El gènere de la persona	Gènere (GEDCOMX)	OPCIONAL.
Facts	Esdeveniments relacionats amb la vida d'una persona.	Llista d'Esdeveniments (GEDCOMX)	OPCIONAL.

Així doncs, podem veure com la instància del recurs Persona conté un camp booleà, que indica si aquesta pot ser utilitzada de forma pública o només en l'àmbit privat i tres camps que es troben codificats sota els estàndards del format GEDCOMX.

Per exemple, el format de dades `http://gedcomx.org/v1/Gender`, representaria un recurs amb l'estructura que s'exposa a la taula 4.2 i els valors possibles per l'enumeració de gènere, s'indiquen en la taula 4.3.

Table 4.2: Codificació GEDCOM X del recurs Gènere

Nom	Descripció	Format de dades	Restriccions
type	Valor que indica el gènere de la persona.	Enumeració	REQUERIT. El gènere ha de ser especificat i es recomana utilitzar un valor dels acceptats per l'enumeració és recomanat.

Table 4.3: Valors enumeració type

Gènere URI (Gedcom X)	Descripció
<code>http://gedcomx.org/Male</code>	Gènere Masculí
<code>http://gedcomx.org/Female</code>	Gènere Femení
<code>http://gedcomx.org/Unknown</code>	Gènere no especificat

Com que l'objectiu del projecte no és estudiar la codificació GEDCOM o GEDCOM X, sinó comprendre quina informació es troba realment disponible a través de l'API de FamilySearch, no entrarem més en detall en aquests formats.

L'objectiu d'aquest apartat era explicar quin és l'estàndard de representació de dades genealògiques utilitzat per l'API. Per qualsevol informació extra que es vulgui consultar, s'adjunta a la bibliografia del projecte, l'enllaç a la documentació del model conceptual.

### 4.3.2 Format de dades FamilySearch

El format de dades FamilySearch, defineix el format d'aquells objectes específics relacionats amb la plataforma de dades pròpia de l'organització. Això implica que

aquestes estructures no formen part de cap estàndard i manquen de sentit fora del context pel qual han estat definides.

L'estructura dels objectes específics de FamilySearch ha estat creada com una extensió de l'especificació GEDCOM X. Per tant, segueixen una estructura molt similar.

### 4.3.3 Format de dades Atom (o Atòmic)

Els formats de dades Atom, o atòmic, és utilitzat per proporcionar un format pel contingut web i les metadades.

Aquest format és utilitzat, entre altres llocs, en les col·leccions ordenades de resultats, com podrien, per exemple, les respostes a la funció de cerca de persones o l'obtenció del historial de canvis d'una persona.

### 4.3.4 Codificacions dels formats de dades

Els formats de dades que s'han exposat en els apartats anteriors no són més que unes convencions que marquen l'estructura a seguir per representar els diferents objectes o recursos utilitzats per l'API de FamilySearch.

Tanmateix, aquestes estructures han de ser codificades per tal de poder ser transmeses a través del núvol i en concret, FamilySearch, proporciona suport a les dues codificacions més comunes i utilitzades per aquesta finalitat. Els llenguatges XML i JSON.

## El llenguatge XML

El XML, és un llenguatge de marcatge que defineix un conjunt de regles a seguir per tal de codificar, documents i informació, en un format llegible i processable, tant per éssers humans, com màquines.

El llenguatge va ser definit pel Consorci World Wide Web i tracta d'emfatitzar la simplicitat, generalitat i usabilitat del model, per l'ús a través d'Internet.

Una versió reduïda de la representació en XML del recurs 'Nota', amb camps: subjecte, text i atribució, quedaria representat de la forma següent:

Listing 4.1: Representació bàsica en XML d'una Nota

```
<Note xmlns='...'>
  <subject>...</subject>
  <text>...</text>
  <attribution id='...'>
    <contributor resourceId='...' resource='...'/>
    <modified>...</modified>
    <changeMessage>...</changeMessage>
    <creator resourceId='...' resource='...'/>
  </attribution>
</Note>
```

```
<created>...</created>
</attribution>
</Note>
```

Com es pot observar, cada camp, objecte o peça d'informació, es troba envoltada per dues etiquetes que en marquen l'inici i final. El text situat a l'interior d'aquestes etiquetes, indica de quin camp es tracta. Per exemple, pel camp subjecte, tenim les etiquetes `<subject>` i `</subject>`.

## El llenguatge JSON

El llenguatge JSON, és un estàndard de format obert que, de la mateixa forma que el llenguatge XML, pretén crear codificacions llegibles tant per éssers humans com màquines i al mateix temps, poder transmetre aquestes dades a través del núvol de forma ordenada.

Aquest format es basa en el concepte 'clau -valor'. És a dir, cada camp d'un objecte a representar està format per una clau i un valor associat a aquesta clau.

El llenguatge JSON deriva del JavaScript. Al principi, només aquesta plataforma incorporava funcions per codificar i decodificar aquest llenguatge de marcatge.

Durant els últims anys, el format JSON s'ha vist convertit en l'estàndard de la indústria per l'intercanvi de dades a través d'Internet i en conseqüència, ha provocat que molts altres llenguatges de programació hagin incorporat les seves pròpies funcions de codificació i decodificació.

Una versió reduïda de la representació en JSON del recurs Nota, amb camps: subjecte, text i atribució, es representaria de la forma següent:

Listing 4.2: Representació bàsica en JSON d'una Nota

```
{
  'lang': '...',
  'subject': '...',
  'text': '...',
  'attribution': {
    'contributor': { },
    'modified': '...',
    'changeMessage': '...',
    'creator': { },
    'created': '...',
    'id': '...'
  },
  'id': '...'
}
```



## 4.4 Evolució temporal de l'API

### 4.4.1 Evolucions al llarg del temps

L'api de FamilySearch s'ha vist subjecte a diversos canvis i modificacions al llarg del temps, algunes, amb més repercussió i implicacions que altres. La llista de canvis és prou llarga com per no intentar reproduir-la en la memòria. De totes maneres, aquesta és adjuntada a la bibliografia del projecte.

L'últim canvi introduït a l'API i de fet, probablement, dels més importants que s'han produït fins ara, fa referència a la nova estructura del back-end anomenada Tree Foundation.

A falta de documentació oficial al respecte, es va consultar un dels Webinair passats, oberts al públic per part de FamilySearch, per tal de comprendre millor la magnitud d'aquests canvis i quines implicacions tenia. En el següent apartat, tractarem més a fons en què consisteix Tree Foundation.

Les primeres fases d'aquest darrer canvi van arribar a producció a mitjans del mes de juliol, moment en què la implementació dels exemples pel projecte a través del SDK oficial de JavaScript estava complerta. Aquests canvis, van repercutir en gran mesura sobre el SDK oficial, trencant-ne moltes de les funcionalitats.

Al final, els principals problemes que va causar pel projecte van ser interrupcions permanents en l'entorn de proves Sandbox i la inhabilitat d'accedir als recursos Font de Dades i Discussió relacionats amb una persona, a causa de la no adaptació del SDK oficial.

### 4.4.2 Tree Foundation

Com s'ha comentat en l'apartat anterior, Tree Foundation és la nova versió del back-end que FamilySearch està implementant. A continuació, es detallen en més profunditat les implicacions d'aquest canvi, ja que pot tenir un gran impacte en com les principals dades emmagatzemades per l'API hauran de ser accedides en un futur no molt llunyà.

El projecte de Tree Foundation consisteix a canviar per complet el motor que fa funcionar el back-end de FamilySearch.

El motor actual, anomenat de forma interna, Conclusion Tree i que fins ara, era l'estructura amb la que l'API es comunicava, caurà en desús en benefici de la nova estructura implementada en el projecte Tree Foundation.

El canvi d'arquitectura té com a raó de ser poder augmentar l'escalabilitat del sistema i fer front a l'elevada demanda de dades sota la que l'API es veu sotmesa. En la imatge 4.1 es pot veure com es passa d'un sistema unificat, a un sistema distribuït.

Els beneficis esperats del canvi són un increment en l'escalabilitat del sistema, una millora eventual del rendiment, tot i que aquest no és l'objectiu principal du-

Figure 4.1: Arquitectura *Conclusion Tree* vs *Tree Foundation*

rant les primeres fases d'implementació i finalment, una millor correlació entre el funcionament del sistema i com l'API accedeix a les dades.

Adicionalment, el projecte canviarà part de la informació inclosa en els recursos accessibles a través de l'API.

Un dels canvis que l'evolució de Tree Foundation pretén integrar i que avui en dia, encara no es troba a producció, és integrar dins del recurs d'una persona tota la informació que s'acostuma a consultar i que ara no contenia. Per exemple, les relacions familiars o les fonts de dades que verifiquen que la informació introduïda és correcta.

Aquest canvi, pretén facilitar la navegació per les dades, evitant haver d'accedir a múltiples recursos per tal d'aconseguir tota la informació desitjada. D'aquesta forma, s'aconseguirà alliberar part la càrrega que suporta el sistema i millorar-ne, per tant, l'eficiència i disponibilitat. La figura 4.2 representa de forma visual el que aquest canvi implica en el recurs Persona.

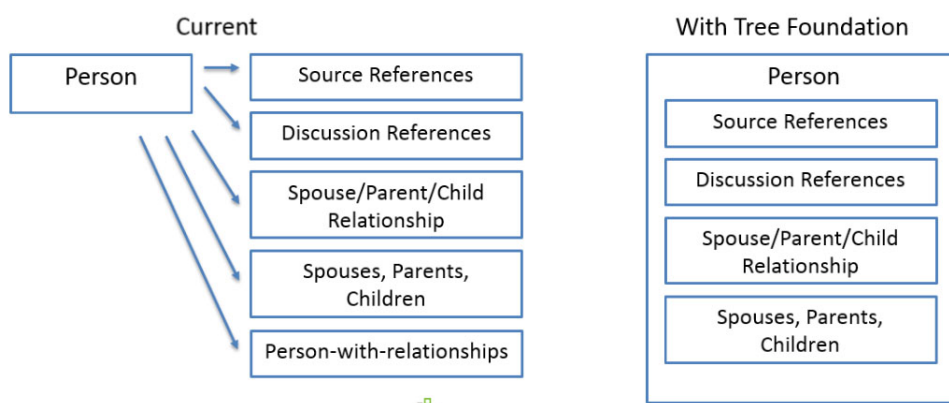


Figure 4.2: Relocalització d'informació sobre alguns recursos.

Quan Tree Foundation va arribar a producció, només els recursos Source i Discussion es trobaven incorporats dins del recurs Persona i els antics enllaços, a aquesta informació quedaven obsolets. Més endavant, apuntant a mitjans d'Agost, s'espera que es realitzi el mateix canvi per les relacions familiars. La figura 4.3 mostra els canvis als quals ens estem referint.

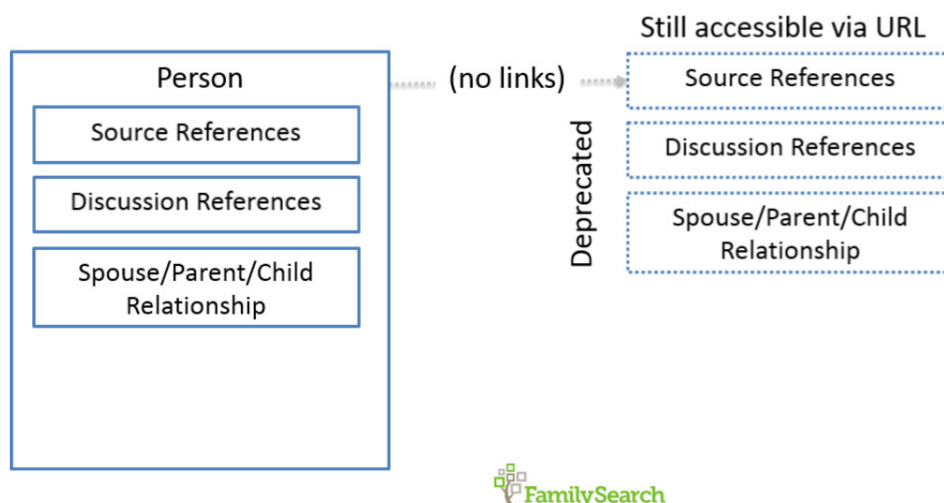


Figure 4.3: Localització dels recursos Discussió i Fonts de dades en la fase 1.

Tot i que el canvi és molt prometedor, cal tenir en compte que té el potencial de trencar part dels SDKs oficials i moltes de les aplicacions que tractin amb la versió antiga de l'API, de la mateixa forma, que ja ho han fet les primeres versions de Tree Foundation arribades a producció.

Per tant, caldrà estar al compte de quan aquests canvis arriben a producció i de quan els SDKs oficials estaran preparats, per ser utilitzats sense error, davant d'aquesta nova versió del back-end de FamilySearch.

## 4.5 Serveis extres oferts per l'API

L'API de FamilySearch no es caracteritza només per ser el centre d'informació obert amb més registres genealògics, sinó que a més a més, l'API ha estat dissenyada i pensada perquè pugui ser utilitzada de forma simultània per moltes persones, des de diferents indrets del món, d'una forma còmode i personalitzada.

Tot i que evidentment cap sistema és infal·lible, i de fet, com ja hem comentat, FamilySearch està realitzant una actualització del seu sistema del back-end per tal de garantir una millor escalabilitat del sistema, el sistema actual ja presenta una sèrie de serveis dignes de menció.

### 4.5.1 Caching

El fet de voler oferir un servei de *caching* decent ha influenciat bona part del disseny de l'API. Davant la pregunta evident de, perquè és important el *caching* a FamilySearch?

- El fet que el client pugui aplicar tècniques d'emmagatzematge, evita la repetició de certes peticions al sistema.
- L'existència de reverse proxies permet, a aquests servidors, respondre a peticions dels clients sense necessitat d'accedir als servidors i bases de dades si detecten que la informació demana, ja es troba disponible en aquests. Per tant, ajuden a reduir la càrrega del sistema.
- Davant noves peticions, els servidors poden indicar als proxies que les seves dades encara són vàlides i que per tant, no fa falta que realitzin una nova petició i n'esperin la resposta.

### 4.5.2 Throttling

La funcionalitat de *throttling* s'encarrega de posar un límit al nombre de peticions, en un cert interval de temps, que un usuari pot realitzar. Aquests límits són creats en l'àmbit d'usuari i per tant, utilitzar més d'una sessió simultània, no permet saltar-se aquesta limitació.

Les polítiques de *throttling* permeses són calculades mitjançant el temps de processat, respecte al temps real que ha transcorregut. Diferents recursos, disposen de diferents polítiques. En cas d'excedir aquests límits, la capçalera de resposta indica quant de temps cal esperar entre peticions, per poder tornar a interactuar amb l'API.

Aquest servei, que evidentment, suposa una limitació pels usuaris, evita que certs usuaris de forma intencionada o involuntària, bloquegin tots els recursos disponibles per part dels servidors i el servei quedi inutilitzable per altres persones.

### 4.5.3 Sincronització

Un dels reptes més comuns, per aquelles aplicacions que emmagatzemen dades d'un tercer, és el de saber si aquestes dades han estat modificades des de l'últim cop que van ser consultades o extretes.

El paràmetre ETag, de la capçalera de les peticions contra l'API, ha estat modificat per part de FamilySearch, per obtenir la versió del recurs consultat. Per tant, si es realitza una petició de només aquestes capçaleres, a canvi d'un petit esforç de l'API, es podria comparar si els recursos desitjats han estat modificats o no sense causar un gran impacte.

Un cop es coneix que un recurs ha estat modificat, es poden descobrir els canvis realitzats mitjançant la comparació dels dos objectes o llegint l'historial de canvis.

Evidentment, sempre es podria realitzar una sobre escriptura completa del recurs.

#### 4.5.4 Internacionalització

Un dels serveis que més m'ha sorprès trobar-me a l'API de FamilySearch és el de suport a la internacionalització.

La internacionalització consisteix a adaptar una peça de software o contingut, a diferents idiomes, sense necessitat de realitzar canvis en el codi. L'API de FamilySearch permet realitzar les peticions de forma que certs conjunts de dades retornades, com poden ser per exemple, el nom de països o persones, estiguin localitzades en l'idioma especificat.

La internacionalització permet, per exemple, que un usuari que es troba a Espanya vegi el terme 'living' (viu), en el seu idioma natiu. Un altre recurs que moltes pàgines web localitzen són les dates. És ben sabut que l'estructura americana d'una data, per exemple, difereix d'una data europea en la posició dels termes any, mes i dia.

El conjunt de dades que FamilySearch permet localitzar són:

- Les propietats de visualització del recurs Persona.
- La informació relativa a les localitzacions.
- Certs aspectes del vocabulari controlat, o comú, de l'API.
- Dates.

#### 4.5.5 Transferència de registres en grans quantitats

Existeix la possibilitat per les organitzacions que així ho desitgin, de signar un acord amb FamilySearch, que els hi permeti mantenir la seva pròpia còpia de dades oficials de FamilySearch, relativa a persones difuntes.

Per realitzar aquestes transferències, que mouen alts volums de dades, FamilySearch disposa d'una API especial que permet a les organitzacions gestionar, mantenir i actualitzar, les dades emmagatzemades en els seus sistemes a través de diferents sessions.

Aquesta opció pot permetre a altres organitzacions ordenar i emmagatzemar les dades amb una estructura diferent i habilitar, d'aquesta forma, la possibilitat de realitzar diferents tipus d'estudi. Per exemple, estudis de mineria de dades.

### 4.6 Eines de desenvolupament

FamilySearch posa a disposició dels desenvolupadors un seguit d'eines destinades a facilitar les seves implementacions contra l'API.

Són dos els recursos principals que permeten apropar la utilització de l'API a qualsevol desenvolupador. Els SDKs i les aplicacions d'exemple.

#### 4.6.1 Els SDK de FamilySearch

Els Kit de Desenvolupament Software (SDK), són un conjunt d'eines que faciliten la creació d'aplicacions per un software, hardware, framework, sistema de computació, videojoc, sistema operatiu o plataforma de desenvolupament determinada.

En el cas de FamilySearch, es tracta d'eines de desenvolupament destinades a un llenguatge de programació específic.

Però, perquè esdevé útil la utilització d'un SDK? En el cas de FamilySearch un SDK ens ofereix la possibilitat de comunicar-nos amb l'API sense haver de preocupar-nos, de forma directa, en com les URIs han de ser codificades o en parcejar els JSON/XML de les respostes. Taques que poden esdevenir tedioses i repetitives.

Per tant, els SDK permeten als desenvolupadors centrar-se en el desenvolupament de funcionalitats i no en la gestió de comunicacions amb l'API.

FamilySearch disposa en l'actualitat de sis SDKs diferents. Tots ells faciliten la interacció amb l'API, encara que no tots permeten les mateixes funcionalitats. A més a més, alguns d'aquests SDK són oficials, significant que són desenvolupats i mantinguts per l'organització FamilySearch, mentre que altres, són creats i gestionats per la comunitat.

Els diferents SDK disponibles es llisten a continuació:

- **Java SDK:** El Java SDK és un SDK oficial que serveix per crear aplicacions d'escriptori amb el llenguatge Java.
- **PHP SDK:** El SDK de PHP és un SDK oficial per crear aplicacions web. Desafortunadament, no es troba del tot actualitzat i l'última versió d'aquest no integra masses de les funcionalitats que es poden realitzar contra l'API.
- **C Sharp SDK:** El SDK de C Sharp és un altre dels SDK oficials i serveix per crear aplicacions amb la tecnologia .NET.
- **Javascript SDK:** El SDK de Javascript és l'últim SDK oficial i està orientat sobretot a la creació d'aplicacions web. És el SDK que s'ha utilitzat en aquest projecte ja que cobreix gran part de les funcionalitats de l'API i en facilita l'ús.
- **Python SDK:** El SDK de Python resultaria molt atractiu de cara a tota mena d'aplicacions, però per desgràcia, encara es troba en fase de desenvolupament i sense dates concretes de finalització.
- **Ruby SDK:** La perla de Ruby és l'últim SDK disponible per FamilySearch i es tracta d'un SDK no oficial.

### 4.6.2 Les aplicacions d'exemple

Les aplicacions d'exemple consisteixen en aplicacions de codi obert que permeten entendre, de forma més pràctica, com funcionen els SDK.

Solen cobrir un conjunt d'operacions bàsiques, com poden ser, per exemple, llegir l'usuari connectat, realitzar una cerca bàsica o la lectura d'una persona concreta de l'arbre familiar. En definitiva, representen un bon punt de partida per aquells que vulguin familiaritzar-se amb un SDK concret i comprendre'n les bases que els permetin l'elaboració de funcionalitats més complexes en el futur.

De la mateixa forma que els SDK, FamilySearch disposa de sis aplicacions d'exemple, aquest cop, sent només oficials les aplicacions que utilitzen els SDK de Javascript i Java.

### 4.6.3 Altres eines intressants

L'últim recurs per desenvolupadors que FamilySearch ofereix, és un llistat d'eines que poden ser interessants o esdevenir útils, de cara a la realització de proves amb l'API.

Entre aquestes, destaquen diferents intermediaris que serveixen per simular peticions HTTP i HTTPS contra l'API de FamilySearch i una utilitat, no oficial, que en teoria permet copiar dades de producció als entorns de desenvolupament.

## 4.7 Procés de certificació

Per tal de poder connectar les aplicacions desenvolupades per tercers al conjunt de dades oficial de FamilySearch, aquestes aplicacions han de ser sotmeses a un procés de certificació.

Les aplicacions poden ser certificades per l'ús comercial o per ús limitat. Les Apps que volen ser certificades només per ús limitat, requereixen un anàlisis tècnic sobre el seu funcionament, mentre que de les aplicacions d'ús comercial també són analitzades des d'un punt de vista de negoci i marketing. Com a contrapartida, les aplicacions d'ús limitat no poden aparèixer a la galeria d'aplicacions.

Hi ha tres tipus de certificacions principals:

- **Certificació de lectura:** L'aplicació ha de ser certificada per la lectura de tots aquells recursos als quals accedeix. També cal que compleixi amb certs estàndards del procés d'identificació.
- **Certificació d'escriptura:** En cas que l'aplicació realitzi operacions d'escriptura, aquestes també hauran de ser revisades per part de FamilySearch.
- **Certificació per transferència d'arxius en grans quantitats:** Per aquelles organitzacions que tinguin permís per accedir als protocols de transferència

de dades en grans quantitats, cal certificar i revisar, conjuntament amb FamilySearch, les operacions realitzades contra aquesta API.

Es veuran més detalls sobre el procés de certificació en la part pràctica de la memòria del projecte.

Un cop una aplicació ha estat certificada, en cas que aquesta modifiqui les operacions de lectura o escriptura que realitza, caldrà tornar a certificar-la abans de desplegar els canvis a producció.

Existeix un procés encarregat de controlar que es compleixen els estàndards marcats per FamilySearch i en cas de no complir-los, pot significar el retirament dels drets d'accés a producció.

Per acabar aquesta secció, comentar que mantenir una relació formar amb FamilySearch, proporciona certs beneficis com aparèixer en les seves galeries d'aplicacions i poder utilitzar el logotip d'aplicació certificada, entre altres petits avantatges.

## 4.8 Diferents entorns de treball

Es disposa de molt poca informació sobre els diferents entorns als quals es pot accedir relatius a l'API de FamilySearch. A força d'investigar diferents codis i possibilitats, s'ha arribat a la conclusió de què existeixen els següents entorns de treball:

- **Sandbox:** Entorn de treball únic per cada usuari. Utilitzable durant el desenvolupament de noves aplicacions i que replica les funcionalitats de l'API sense tenir accés a les dades de producció. Es pot entendre com un entorn de proves.
- **Staging:** Entorn de treball que s'utilitza per validar que una aplicació, que s'ha desenvolupat en un sandbox, està realment preparada per accedir a les dades de producció.
- **Beta:** Entorn de treball sobre el qual l'API va desplegant noves versions. Quan es vol realitzar un canvi en les funcionalitats o estructura de l'API, s'utilitza aquest entorn perquè els desenvolupadors puguin testear el funcionament de les seves aplicacions abans de desplegar la nova versió a producció.
- **Producció:** Entorn de treball que pot accedir a les dades oficials de FamilySearch. Totes les aplicacions aspiren a connectar-se a aquest entorn de treball.



## Secció 5

# Valoració final sobre la potencialitat de l'API

### 5.1 Introducció

Aquesta secció de la memòria pretén cobrir, des del meu punt de vista personal basat tant en el coneixement adquirit mitjançant l'estudi teòric de l'API, com en les petites pinzellades tècniques que s'han pogut aprendre durant la implementació dels exemples, el potencial que emmascara aquesta API de cara a generar propostes de projecte per futurs estudiants.

En conseqüència, a pesar de la localització dins de la memòria d'aquest apartat, aquest és redactat després d'adquirir el coneixement pràctic bàsic, gràcies a la implementació dels exemples, que seran exposats en les següents seccions de la memòria.

Per comprendre la potencialitat global d'aquesta API, cal dividir-ne l'estudi en diferents blocs o peces. Intentar emetre un judici de valor global, sense sintetitzar-ne primer diferents oportunitats i complicacions de conceptes més específics, no aconseguiria transmetre la profunditat i complexitat de l'abast d'aquest pregunta.

### 5.2 Distribució geogràfica de les dades

Per comprendre la potencialitat global d'aquesta API, cal dividir-ne l'estudi en diferents blocs o peces. Intentar emetre un judici de valor global, sense sintetitzar-ne primer diferents oportunitats i complicacions de conceptes més específics, no aconseguiria transmetre la profunditat i complexitat de l'abast d'aquest pregunta.

FamilySearch, pot presumir de tenir una de les bases de dades d'informació genealògica oberta al públic més gran del món, si no la més gran, amb més de quatre bilions de registres.

A pesar que el nombre de 4 bilions pugui semblar molt elevat, si el comparem amb

els més de 33 bilions de persones que han nascut aproximadament des del 1200 fins a l'any 2011 (agafant així, una dada de referència pública que s'ajusti més o menys al període de temps sobre el que FamilySearch disposa d'informació), ens adonem del fet que disposem d'una mostra acceptable, però lluny de suposar una representació real.

Cal també tenir en compte que els 4 bilions de registres emmagatzemats a FamilySearch no es troben repartits de forma proporcional sobre les diferents regions o països, sinó que l'organització, de forma evident, disposa més dades en aquells indrets en què històricament ha tingut més presència o facilitat d'accés a dades.

D'aquesta forma, els Estats Units d'Amèrica, amb 991 col·leccions de dades diferents, ofereix la informació de 2,5 bilions de registres que daten entre els anys 1500 i 2015. En altres paraules, un 62% del volum total de dades.

De forma paral·lela, i per oferir una escala diferent, Espanya, amb 45 col·leccions, ofereix la informació de 24 milions de registres compresos entre els anys 1251 i 2013. Per tant, resulta fàcil observar que la dispersió de les dades i volum difereix molt segons la regió que vol ser consultada.

Cal doncs, tenir en compte aquesta limitació de cara a proposar o realitzar certa mena de projectes, com poden ser per exemple, els estudis estadístics d'aspectes demogràfics. Com a recomanació, s'encoratja als estudiants a utilitzar aquelles regions, com els Estats Units, més plegades de registres, de cara a funcionalitats generals.

### 5.3 Dades contemporànies

Un dels inconvenients de les dades genealògiques és que aquestes generalment es troben subjectes a lleis de protecció, durant períodes de temps prolongats, abans de poder fer-se públiques. És més, en casos especials com els que hem esmentat en les primeres seccions de la memòria, aquestes poden inclús no arribar mai al domini públic.

Aquest aspecte implica que el valor percebut, del conjunt de dades disponible a través de FamilySearch, sigui més elevat en projectes enfocats al passat, que no pas estudis més contemporanis.

La ironia en aquest punt de la memòria és que a menys que les legislacions canviïn per complet, l'afirmació de què sempre es disposarà de menys dades contemporànies, seguirà sent aplicable independentment dels anys que passin.

Un altre fet que pot impactar a la quantitat de registres contemporanis disponibles és la quantitat d'afiliacions a l'Església de Jesucrist dels Sants dels Darrers Dies i és que cal no oblidar, que aquesta organització, representa al principal benefactor de FamilySearch i per tant, principal origen de font de dades.

## 5.4 Recursos i funcionalitats

El conjunt de recursos utilitzables i les funcionalitats creades al seu voltant, esdevenen un dels punts més favorables de l'API.

El conjunt de paràmetres accessibles relacionats a una persona, o qualsevol altre recurs, resulta immens. Des dels esdeveniments principals relacionats a la seva vida d'una persona, fins a petits detalls com els diferents noms que la persona va rebre al llarg de la seva vida. La informació es troba molt ben estructurada i tots elements relacionats, resulten fàcilment accessibles.

La robustesa de les dades tampoc és cap broma, mitjançant el sistema de canvis es pot desfer qualsevol ús malintencionat o involuntari sobre el conjunt de dades. Clarament, FamilySearch es pren molt seriosament poder garantir la qualitat de les dades.

Per si tot el conjunt de recursos accessible no fos suficient, FamilySearch posa a la disposició dels usuaris una sèrie de funcions de conveniència que permeten, entre altres exemples, cercar persones duplicades, accedir a les ascendències i descendències d'una persona de forma reglada i estructurada i delimitar les cerques per més paràmetres dels que un es podria imaginar.

Tot plegat, el conjunt de recursos, granularitat de la informació i les funcionalitats de fàcil accés, converteixen l'API de FamilySearch en un poderós aliat de cara a la recerca genealògica.

## 5.5 Naturalesa de l'API

Un dels primers xocs que em vaig emportar quan vaig començar a estudiar més a fons la potencialitat de l'API, va ser perquè fins aquell moment no havia tingut en compte el motiu pel qual aquesta havia estat concebuda.

L'API de FamilySearch neix per ajudar a individus particulars a realitzar recerca genealògica, sobre els seus avantpassats o els d'un tercer. Cal tenir molt present aquesta definició, ja que limita o debilita en gran mesura, els possibles projectes a realitzar.

FamilySearch va dissenyar l'API perquè un usuari pogués realitzar una cerca, de la forma més específica possible, després de la recopilació prèvia d'informació sobre la persona cercada. Per tant, no està pensada per accedir a un gran volum de registres de forma simultània, accedir a les dades per un nivell de granularitat inferior a la del concepte 'persona', ni realitzar moltes peticions consecutives contra la plataforma.

En conseqüència tota aspiració de realitzar projectes de mineria de dades o estudis de baixa granularitat, queden completament descartats, o si més no, condicionats en gran mesura de cara a la implementació tècnica o automatització de tasques.

## 5.6 Utilització de l'API en el marc d'un PFC

Un últim concepte a destacar és com encaixa la utilització d'aquesta API en el marc d'un projecte final de carrera. És a dir, si deixem de banda la discussió de com és de potent aquesta, resulta factible utilitzar-la de cara a un projecte final de carrera?

Crec que resulta de vital importància respondre a aquesta pregunta de forma independent a la de la potencialitat, per no vincular dos conceptes diferents.

Un projecte final de carrera es desenvolupa generalment en el període de temps equivalent al d'un quadrimestre. Com ja s'ha esmentat en altres seccions de la memòria i encara tornarà a aparèixer més endavant, per tal d'aconseguir accés a les dades de producció de FamilySearch, cal certificar l'aplicació.

Aquest procés, com aquest projecte n'és una clara mostra, pot resultar complicat i ple de complicacions, i encara podria esdevenir més complex si es volgués realitzar una aplicació amb drets d'escriptura a l'arbre genealògic o comercialitzable.

El que volem indicar en aquest apartat és que si la planificació del projecte no és bona, i inclús així, s'incorre en un cert risc, existeix la possibilitat de no disposar del temps suficient per implementar, certificar i extreure les conclusions necessàries, sobre les dades de producció.

El fet que existeixi en l'actualitat, la possibilitat de certificar les aplicacions per ús personal i no només comercial, augmenta en gran mesura les possibilitats dels estudiants a aventurar-se en projectes de recerca genealògica.

En certa forma, esdevé probable que la utilització d'aquesta API condicioni l'estructura dels projectes de la mateixa forma que ho ha fet amb aquest. Forçant un clar esforç inicial per acabar la implementació tan aviat com es pugui, per tal de poder experimentar amb les dades de producció, o en el nostre cas, esbrinar de què es tractava exactament aquest procés de certificació.

Això no obstant, s'espera que l'estudi realitzat en aquest projecte representi una facilitació i acceleració considerable de la corba d'aprenentatge pels futurs estudiants, el que els permetria accedir abans a producció.

En resum, si, l'API de FamilySearch és un recurs utilitzable de cara a la realització de projectes finals de carrera, però cal tenir en compte les seves peculiaritats de cara a la planificació i pot resultar més atractiu a aquelles persones que pretenguin tenir clar, el projecte que volen realitzar en profunditat, abans de matricular-lo o inclús començar-lo amb un quadrimestre d'antelació, al que serà matriculat.

## 5.7 Conclusió

Considerades les limitacions geogràfiques, temporals, estructurals i temporals (en l'àmbit de temps del que es disposa per realitzar un projecte) podria semblar que no queden moltes opcions possibles, de cara a plantejar propostes de projecte, més

enllà de la recerca genealògica bàsica. Malgrat això, no és la meua opinió exacta que aquest en sigui el cas.

Si bé és cert, que cal pensar en propostes de projecte que encaixin dins del marc delimitat per aquestes restriccions, les eines posades a disposició dels usuaris i la quantitat d'informació disponible, permeten la utilització de vies secundàries per tal d'assolir diferents objectius i ajudar així a respondre certes preguntes que aquest projecte no ha tingut temps d'explorar.

En la següent secció de la memòria es podran observar un conjunt de propostes de projecte que encaixen dins del marc descrit en aquesta secció i que pretenen oferir resposta, a certes preguntes més específiques, sobre la potencialitat o possibilitats d'ús d'aquesta.



## Secció 6

# Llista de propostes de projecte

### 6.1 Introducció

Aquesta secció recopila un conjunt d'idees que poden servir com a projectes finals de carrera per estudiants de la Facultat d'Informàtica de Barcelona.

L'objectiu de cada una de les propostes no és la de representar un enunciat tancat, sinó oferir pistes sobre diferents implementacions possibles que puguin inspirar als estudiants a modificar-les, combinar-les, reduir-les, ampliar-les o crear-ne de noves.

Com es podrà veure, moltes d'aquestes propostes giren al voltant d'estudis històrics o la validació de les dades emmagatzemades en els sistemes de FamilySearch, respecta la realitat. El motiu, és que més enllà de la funcionalitat de cerca, la principal preocupació sobre aquestes dades és amb quin grau d'exactitud representen la realitat que les envolta.

A més a més, recordar que l'API de FamilySearch es troba sempre en constant evolució, i que per tant, és una bona idea revisar la viabilitat de cada proposta abans de decidir, amb total certesa, el projecte que es vol realitzar.

### 6.2 Comparació sobre la popularitat de noms

L'objectiu d'aquesta funcionalitat és comparar la popularitat d'un o més noms, en un període concret del temps i amb la possibilitat de fixar la regió demogràfica a consultar.

La idea principal és que donada la introducció d'un o més noms, el sistema cerqui el nombre d'instàncies de persones nascudes amb el nom especificat, en els deu anys anteriors o posteriors a la data indicada, per la regió geogràfica especificada.

D'aquesta forma, es podria comparar quin dels dos noms ha estat més popular, segons les dades de FamilySearch, any a any.

Al mateix temps, esdevé interessant permetre la cerca d'un sol nom per observar

si certs esdeveniments històrics han pogut influenciar el nombre de nadons amb un cert nom. Per exemple, suposa l'elecció d'Obama com a president dels Estats Units, un increment en el nombre de persones nascudes amb aquest nom durant els següents anys?

La segona raó de ser de l'eina és ajudar a decidir, per exemple, el nom dels fills d'una persona, comparant, d'aquesta forma, la popularitat actual dels noms que s'estiguin avaluant.

A continuació llistem diferents possibilitats d'extensió:

- Ampliar la cerca a diferents països, on per cada país, el nom introduït serà localitzat. Per exemple, si l'usuari introdueix Alexander, la cerca a Espanya fos realitzada amb el nom d'Alejandro o Alex. En cas de no trobar aquesta base de dades, sempre es podria crear una taula manual d'exemple, amb unes quantes llengües i noms i utilitzar-la.
- Comparar instàncies de noms a Catalunya extrets de les dades de FamilySearch, amb la comparació real extreta del institut nacional d'estadística.

### 6.3 Portal de cerca localitzat al Català

En aquesta memòria ja hem parlat de la funcionalitat de localització habilitada per part de FamilySearch i encara que aquesta suporta la localització a la llengua espanyola, no ho fa per la catalana.

La idea d'aquesta funcionalitat és oferir a l'usuari un portal de cerca en català sobre les dades de FamilySearch, que no només faciliti la comprensió de la cerca a aquelles persones que vulguin utilitzar el català, sinó que també localitzi, en la mesura que sigui possible, la resposta retornada per l'API.

Per exemple, es podria localitzar la informació relativa a l'estat actual d'una persona: living o deceased, que podria ser mostrada com a viva o difunta. De la mateixa forma, es pot localitzar tot el contingut de la resposta, des del nom dels camps d'informació, al contingut d'aquests en algunes situacions.

Alguns exemples de possibles extensions pel projecte són:

- Restringir la cerca del portal a només Catalunya, amb la possibilitat de desactivar la funció. A més a més, oferir ajuda de refinació a la cerca, de cara a introduir les diferents províncies o ciutats, evitant que l'usuari introdueixi valors invàlids.
- Posar-se en contacte amb l'organització FamilySearch per convertir la localització realitzada a la llengua Catalana, en una localització oficial acceptada pel sistema.



## 6.4 Geolocalització d'un cognom en diferents nivells

Aquesta funcionalitat pretén ampliar l'exemple programat en aquest projecte final de carrera, evolució geogràfica d'un cognom.

L'exemple implementat només permet la visualització d'instàncies d'un cognom al nivell de país. Aquesta nova funcionalitat hauria de permetre, com a mínim, les visualitzacions en l'àmbit de continent i en l'àmbit d'estat o comunitat autònoma. De totes maneres, com més nivells de profunditat diferents poguessin ser utilitzats, millor.

L'usuari haurà de ser capaç de navegar per aquests diferents nivells de forma interactiva mitjançant el mapa (existeixen mapes interactius a disposició dels desenvolupadors) o controls específics.

La dificultat d'aquesta eina, a diferència de la implementada com a exemple, és que o bé farien falta més peticions contra l'API o realitzar una petició més gran i explorar després els registres d'un a un. Una alternativa, podria ser realitzar les peticions a l'API, després que l'usuari indiqués que vol canviar el nivell mostrat i anar emmagatzemant la informació de forma local.

## 6.5 La recerca genealògica i l'heràldica

Aquesta funcionalitat vol relacionar les dues ciències que sempre han caminat de la mà, la genealogia i l'heràldica.

El primer objectiu del projecte implicaria identificar quina font de dades podria ser utilitzada per obtenir els diferents escuts d'armes. Probablement, no existeixi cap API en línia de la qual es puguin obtenir i caldrà realitzar una extracció automàtica d'alguna pàgina web o aplicació d'escriptori.

Un cop es disposin dels diferents escuts d'armes, es proposa realitzar una implementació de cerca simple per veure, conjuntament amb els detalls d'una persona, l'escut d'armes del nom de família.

Un cop es disposin dels diferents escuts d'armes, es proposa realitzar una implementació de cerca simple per veure, conjuntament amb els detalls d'una persona, l'escut d'armes del nom de família.

Una possibilitat d'extensió, que podria ser bastant atractiva, consisteix a implementar una API que serveixi els diferents escuts d'armes i que la nostra aplicació l'utilitzi per complementar les dades de les persones trobades a l'API de Family-Search.

## 6.6 Projectes d'indexació

Tot i que aquesta proposta no pretén interactuar directament amb l'API de FamilySearch, volíem realitzar com a mínim una proposta que estigués relacionada amb el procés d'indexació.

Existeixen dos processos d'indexació diferents, els que es realitzen sobre fitxers amb un format específic i els que es basen en la transcripció d'imatges a través del software de FamilySearch. Aquesta proposta de projecte, és en realitat dividida, en dues diferents.

La primera, aconseguir accés a algun registre genealògic local o posar-se amb contacte amb alguna organització que vulgui pujar el contingut de registres amb un format específic, al núvol. Sobre aquest registre, implementar un sistema d'automatització que transcriu les dades i les prepari per ser enviades a FamilySearch.

La segona possibilitat, és realitzar un programa que interactuï amb les imatges digitalitzades, llegeixi les seccions de la imatge sobre les que s'ha d'extreure la informació i intenti informar a l'usuari, que no omplir automàticament, sobre el contingut dels camps.

La gràcia d'aquest projecte és que es podria realitzar sense preocupar-nos per la certificació de l'aplicació, ja que per indexar registres un només s'ha de declarar com a voluntari i començar a experimentar.

Aquestes dues propostes esdevenen, amb una alta probabilitat, bastant complexes, per aquest motiu, es prega a l'estudiant que realitzi un bon estudi previ sobre l'abast i viabilitat del que vol realitzar abans d'inscriure el projecte.

## 6.7 La història de l'església mormona a través de FamilySearch

En una de les primeres seccions de la memòria, s'ha exposat per sobre la història de l'església mormona. Com s'ha pogut observar, aquesta ha estat marcada per nombroses expulsions de diferents territoris i conflictes.

L'objectiu d'aquesta funcionalitat seria intentar traçar una relació entre els diferents emplaçaments o seus principals del col·lectiu, en els diferents moments del temps i les dades emmagatzemades per FamilySearch.

És a dir, destaca FamilySearch per tenir dades sobretot d'aquelles zones en les quals s'ha trobat especialment pressent en comparació a la resta de localitzacions? Com d'evident és aquesta relació, si és que existeix? Podem deduir aleshores que la mostra de les dades no és representativa sota cap circumstància?

Aquestes són algunes de les preguntes que aquest projecte podria intentar respondre.

## 6.8 Diversitat geogràfica d'un cognom: Els nostres avant-passats

Aquesta funcionalitat ha estat inspirada per una campanya publicitària creada per l'agència de viatges Momondo, anomenada, *The DNA Journey*. Recomanem la visualització del vídeo<sup>1</sup> per comprendre millor les motivacions darrere d'aquesta funcionalitat.

Totes les persones creiem que som bastant autòctones del lloc on hem nascut, però els estudis d'ADN sobre els nostres gens ens poden deixar molt sorpresos i demostrar que gairebé tothom té una diversitat ètnica i geogràfica important en els seus gens.

Evidentment, no tothom vol pagar per un test d'ADN i la funcionalitat que proposem, tot i que evidentment, no podrà comparar-se amb aquesta opció, pretén intentar informar a les persones que el seu propi cognom apareix en tota mena de països.

L'objectiu d'aquesta funcionalitat, de forma similar a l'exemple implementat sobre l'evolució geogràfica d'un cognom, és trobar una forma eficient d'oferir a l'usuari un llistat del nombre total o percentatge d'instàncies d'un cognom, per cada país i en diferents èpoques. Podria ser una bona idea, restringir la cerca als primers deu o vint països més grans de cada continent.

El repte del projecte, no és només la consulta de les dades, sinó trobar una forma intel·ligent d'extrapolar l'influència d'un cognom en una regió determinada, en comptes de basar-nos en el total d'instàncies com ha fet l'exemple implementat, el que provoca que el resultat estigui condicionat pel nombre de registres disponibles en cada país.

Queda per tant, a decisió de l'usuari, com realitzar una interpretació de la influència d'un cognom en cada regió i evidentment, sempre de forma aproximada. L'objectiu final és fer palpable la diversitat geogràfica dels nostres possibles avant-passats o relatius.

Aquesta proposta pot ser complicada degut al control sobre el temps d'execució. En cas de no veure viable realitzar una proposta que extregui les dades en temps real, sempre es pot realitzar un estudi específic sobre alguns cognoms d'interès o persones conegudes i mostrar-ne els resultats.

## 6.9 Les col·leccions de dades de FamilySearch

L'objectiu d'aquesta aplicació és llegir les diferents col·leccions o fonts de dades de FamilySearch i llistar-ne, de quines regions i sobre quins períodes de temps, contenen registres genealògics.

---

<sup>1</sup><https://www.youtube.com/watch?v=tyaEQEmt5ls>

L'objectiu, és oferir als usuaris un portal d'informació que permeti, mitjançant la introducció d'un país, període de temps o ambdues condicions, visualitzar les diferents col·leccions disponibles i la informació específica d'aquestes.

Per exemple, donada la introducció del país Espanya i el període 1500–2000, la funcionalitat hauria de llistar totes les col·leccions que contenen dades que compleixen aquestes condicions, quants registres totals suposen aquestes, quants d'aquests estan accessibles a través de FamilySearch, quans pendents d'indexar, etcètera, etcètera.

Aquesta funcionalitat pretén respondre a un dels problemes principals de l'API i és la falta d'informació sobre la informació emmagatzemada. D'aquesta forma, mitjançant un petit anàlisi previ, podríem esbrinar si la informació que desitgem és probable que existeixi o no, sense perdre el temps en l'exploració manual de registres.

## 6.10 FamilySearch i la segona guerra mundial: Natalitat i Defuncions

Durant el transcurs del temps han succeït un gran nombre d'esdeveniments que han afectat a la població mundial de diferents formes. Un dels conflictes que ha causat més repertori, ha estat la segona guerra mundial.

L'objectiu d'aquesta proposta de projecte és observar l'impacte que va tenir aquest esdeveniment en l'índex de natalitat i defuncions, dels diferents països implicats, a través dels anys del conflicte. Es recomana ampliar la finestra de temps estudiat més enllà dels anys del conflicte per observar quins eren els valors normals, previs i posteriors, a l'esdeveniment.

L'estudiant haurà de trobar la forma d'escalar les dades de cada país segons el volum de registres disponibles.

Una altra tasca que pot realitzar l'estudiant, és comparar els valors obtinguts a través de FamilySearch amb les dades oficials de la segona guerra mundial i respondre preguntes de l'estil: Es corresponen els països amb un increment de defuncions més elevat amb els que van patir més durant la segona guerra mundial?

## 6.11 FamilySearch i la segona guerra mundial: Increment en els casaments

Un estudi realitzat per Randal S. Olson<sup>2</sup>, demostra que quan la segona guerra mundial va esclatar als Estats Units, es va registrar un increment enorme del nombre de casaments al llarg del país., demostra que quan la segona guerra mundial va esclatar als Estats Units, es va registrar un increment enorme del nombre de casaments al llarg del país.

---

<sup>2</sup><http://www.randalolson.com/2015/06/15/144-years-of-marriage-and-divorce-in-1-chart/>

## 6.12. FAMILYSEARCH I LA SEGONA GUERRA MUNDIAL: LA LLISTA DE SCHINDLER61

L'objectiu d'aquesta funcionalitat és intentar replicar aquest estudi pels estats units i estendre'l als països més implicats en la segona guerra mundial, per veure si l'afecte va ser el mateix en diferents regions del globus terraqui.

### 6.12 FamilySearch i la segona guerra mundial: La llista de Schindler

Aquesta proposta pretén realitzar un estudi sobre un dels col·lectius que es va veure més afectat durant la segona guerra mundial, els jueus.

Aquesta proposta de projecte ofereix a l'estudiant parcejar els cognoms coneguts d'aquelles persones que van formar la llista de Schindler, i realitzar un estudi d'aquests sobre les dades de FamilySearch.

El projecte pot intentar respondre preguntes com: Van disminuir en gran quantitat el nombre de registres amb els cognoms indicats? Van realitzar aquestes persones un procés d'emigració a diferents indrets del món? Quina mena de documents enregistrats han quedat d'aquestes persones?

La proposta que ens ocupa se'm va acudir quan vaig descobrir, en les fases prèvies del projecte, el certificat d'emigració de Wladyslaw Szpilman, també conegut, com el pianista de Varsòvia. La imatge 6.1 mostra el registre físic trobat a FamilySearch.

REPÚBLICA DOS ESTADOS UNIDOS DO BRASIL  
FICHA CONSULAR DE QUALIFICAÇÃO  
MODELO S.C. 139

Esta ficha, expedida em duas vias, será entregue à Polícia Marítima e à Imigração no porto de destino

Nome por extenso Wladyslaw Szpilman  
Admitido em território nacional em caráter TRANSITO: Polônia  
(temporário ou permanente)  
Nos termos do art. 6 letra ..... do dec. n. 7967, de 1945  
Lugar e data de nascimento Sosnowiec, 5.12.1911 /  
Nacionalidade polonesa Estado civil casado  
Filiação (nome do Pai e da Mãe) Stanislaw e Eduarda  
Profissão pianista  
Residência no país de origem Hotel Dorá, nesta  
NOME IDADE SEXO

FILHOS MENORES DE 18 ANOS {  
SELO CONS

Passaporte CA0028462 expedido pelas autoridades de Polícia de Varsovia  
na data 25.3.1957  
visado sob n. 8537  
Assinatura do portador: Oswaldo Biato  
Consulado Geral do Brasil em Buenos Aires  
PELO CONSUL GERAL 1º AGO 1957  
PELO CONSUL: Geral  
OSWALDO BIATO  
VICE-CONSUL

NOTA—Esta ficha deve ser preenchida à máquina pela autoridade consular, sendo as duas vias em original.

Figure 6.1: Registre d'emigració de *Wladyslaw Szpilman*

### 6.13 La gran recessió o altres esdeveniments històrics

L'exemple específic que es proposa per aquesta funcionalitat, s'aprofita del fet que es disposa d'un major nombre de registres pels Estats Units que no per la resta de països. Malgrat això, aquesta pot intentar ser reproduïble mitjançant qualsevol altre esdeveniment històric.

L'objectiu de l'exemple proposat és estudiar els impactes de la gran recessió en la població dels Estats Units a través de les dades de FamilySearch i respondre a la pregunta de si la realitat observada és similar als fets reals.

Es suggereix a l'estudiant que estudiï el nombre de morts i emigracions al voltant d'aquest període com a punt de partida.

També volem aprofitar aquesta proposta per suggerir als estudiants que qualsevol fet històric, amb un cert impacte, pot intentar ser estudiat a través d'aquesta API. Per aquest motiu, es convida als estudiants a plantejar els seus propis casos d'estudi.

### 6.14 Comparacions amb els amics o seguidors de Facebook i Twitter

L'objectiu d'aquesta funcionalitat és proporcionar a l'usuari, mitjançant les API de FamilySearch, Facebook o Twitter, un conjunt d'eines per tal de comparar les seves dades amb les dels seus amics.

El nombre de comparacions o eines de cerca a FamilySearch que podrien ser utilitzades és il·limitada, però per donar alguna idea als estudiants del que tenim al cap, en citem algunes a continuació:

- Quin dels nostres amics o seguidors té el nom o cognom més popular en el país on viuen o arreu del món?
- Quin dels nostres amics és més probable que tingui ascendència als estats units (per instàncies del cognom en el país)?
- Quina diversitat cultural representen les teves amistats?
- Mostrar l'esperança de vida mitjana de les persones enregistrades a FamilySearch amb el mateix nom que algun dels nostres amics i comparacions amb aquestes dades.

Les possibilitats són realment il·limitades i a més a més, l'aplicació podria permetre accions com, seleccionar només els amics que ens interessin per certes consultes, realitzar una cerca sobre tots els nostres amics per localitzar de forma fàcil als que ens interessin, etcètera.

## 6.15 Comparació de dades genealògiques reals amb FamilySearch

Aquesta proposta de projecte pretén validar, en certa forma, com de bé representen les dades de FamilySearch la realitat d'un país o països a través de diferents èpoques, o per una època determinada.

Aquesta proposta podria ser dividida en molts projectes diferents, un per cada país, concepte o època que l'estudiant vulgui explorar. Per ajudar a comprendre als estudiants al que ens estem referint, a continuació citem una sèrie d'exemples:

- Comparació de mortalitat per infants: Global, per continents, països, èpoques, etcètera.
- Mitja de fills per família: Global, per continents, països, èpoques, etcètera.
- Edat mitjana de totes les persones enregistrades, en un moment i localització determinades? Evolució d'aquest indicador al llarg del temps.
- Proporció d'homes i dones: Global, per continents, països, èpoques, etcètera.
- Esperança de vida per les dones i homes: Global, per continents, països, èpoques, etcètera.

Aquesta proposta no pretén que l'estudiant abordi tots els conceptes diferents que es pugui imaginar, però si en aquells que cregui que poden tenir un valor més elevat de cara a comparar regions i èpoques.

Segons els factors a estudiar, la complexitat de com hauran de ser processades les dades canvia, i per tant, caldrà tenir-ho en compte a l'hora de definir l'abast del projecte.

Es recomana també als estudiants que cerquin estudis estadístics sobre la població del món, per poder inspirar-se de cara al plantejament de propostes de projecte. A nosaltres ens va ajudar la presentació a les conferències Ted de Kim Preshoff<sup>3</sup>.

## 6.16 Estudi de profunditat dels arbres familiars

Aquesta funcionalitat es basa en l'estudi de les genealogies disponibles a través de FamilySearch.

Es planteja a l'usuari un estudi de la profunditat d'aquestes, quantes generacions diferents estan emmagatzemades de mitjana, com de completes solen estar les dues bandes de l'arbre familiar, etcètera.

Un segon conjunt de dades que l'estudi pot intentar respondre és quins són els períodes de temps en els que era més probable mantenir un arbre genealògic. S'està

---

<sup>3</sup><https://www.youtube.com/watch?v=RLmKfXwWQtE>

perdent la tradició? Ha anat en augment durant els últims anys? Quina mena d'informació és més probable que es trobi disponible?

Aquests són alguns dels exemples que plantegem als futurs estudiants.

## 6.17 Algoritme de marcatge de duplicats

FamilySearch té una funció que permet, donada una persona, aconseguir les persones de l'arbre que tenen una alta probabilitat de ser un duplicat.

L'objectiu d'aquesta aplicació seria realitzar un algoritme, que donada una persona, estudies les persones marcades com a candidates a ser un duplicat i avalués si aquestes marques tenen pinta de ser correctes o no.

L'estudiant podria cercar i comparar segons la diferent informació disponible de cada persona i emetre una conclusió final de diferents nivells, com per exemple:

- Insuficient informació per concloure.
- Duplicat descartat per inconsistència en els naixements.
- Duplicat real amb coincidències de dates de naixement.
- Etcètera.

Un altre aspecte que el projecte podria intentar atacar és comparar la fiabilitat d'aquest algoritme amb la identificació de duplicats per part de FamilySearch.

Aquest projecte pot resultar bastant complex, i a priori, es desconeix la precisió o condicions sota les quals FamilySearch marca a una persona com a duplicada. Per tant, es recomana realitzar un bon estudi previ d'aquests conceptes abans d'embarcar-se en el projecte.



