

# GEDCOM

als Format für die  
Langzeitarchivierung  
von genealogischen Daten

?



Diedrich Hesmer, Herrenberg (D) / Savièse (CH)  
ofb.hesmer@bluewin.ch - <http://ofb.hesmer.name>

## Agenda

- Was ist GEDCOM
- Warum benötigt man dies
- Eignung zur Archivierung
  
- Wie funktioniert GEDCOM
- Programme & GEDCOM
- GEDCOM Flexibilität → Problemstellung
  
- Import von GEDCOM Dateien
- Archivierung → GEDCOM Export – was ist zu berücksichtigen
- Literatur



## GEDCOM – was ist das

- **GE**nealogical **D**ata **CO**munication – Code der Computergenealogie
  - Internationaler Standard zur Beschreibung + zum Austausch genealogischer Daten
  - wird teilweise von verschiedenen Programmen unterschiedlich genutzt und interpretiert auch abhängig von deren Funktionsumfang
  - erlaubt "Nutzer-definierte" Tags → in fast jedem Programm enthalten
  - bietet Vielzahl von Umschreibungen, z.B. beim Datum → aber nicht jeder hält sich daran
  - erlaubt Austausch zwischen Programmen → mit Einschränkung
- Heutiger aktueller Standard : GEDCOM 5.5 / 5.5.1 Entwurf / 5.5 EL
  - ca. 140 Standard Tags + beliebige "Nutzer-definierte" Tags (> 250)
- Entwicklung → Ursprung bei "Kirche Jesus Christ" – LDS

1.0	2.0	...	5.0	5.5	5.5.1	5.5 EL	X
1984	12.85		12.91	12.95	10.99	10.04	02.12
	ANSEL ASCII PAF 2.0			Unicode	UTF-8 Technolog Medien	DE Ortsver- waltung	Initiative RootsTech div AGs



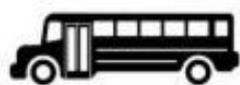
## GEDCOM – warum benötigt man dies (I)

- Datenaustausch zwischen Programmen für ...
  - Erfassungs-Programm zu Auswerte-Programm "A" → "B"
  - Programm-Wechsel "A" → "C"
  - Gegenseitiger Austausch zwischen Kollegen "A" ↔ "D"
- Speicherung der (eigenen) Genealogie-Daten
- ged-Dateien entsprechen nicht immer dem Standard
- Datenverluste häufig unvermeidbar
  - 
- Gute Programme sollten für unbekannte oder nicht implementierbare Tags und deren Daten beim Einlesen der Datei
  - den Anwender fragen
  - diese in einer **Logdatei** erfassen
  - optional in Notiz-Felder ablegen



## GEDCOM – warum benötigt man dies (2)

### Abhilfe

- Deutschsprachiges Autoren-Team von 24 Programmen hat Vereinbarungen für Export/Import von ged-Dateien erstellt (2009 ..13)
  - Basis: GEDCOM 5.5.1
  - Lösung von Defiziten des Standards
  - Einheitliche Interpretation des Standards
  - Eindeutige Vereinbarungen von "Nutzer-definierten" Kennzeichen (Tags)
- Nutzung von Tools zur Korrektur von „fehlerhaften“ Daten im Programm / der ged-Datei
- Trotz allem ist ein Verlust nicht immer zu vermeiden
  -  →  →  →  → ... Kapazität
  - Eingabe Felder (Anzahl, Textformate, Textlängen, ...)
  - Strukturen (eingebettete Texte vs. eigenständige Datensätze)



### GEDCOM – zur dauerhaften Archivierung ?

**!! JA !!**

- Es ist nur ein Format, kein Speichermedium
  - Reine Textdatei → mit jedem Texteditor zu öffnen
  - Standardisierter Aufbau → Stufen Nrn, Befehls Worte (Tags), Texte
  - Sollte auf Computer-lesbare Medien gespeichert werden
- Es ist der einzige Standard
  - Einzelne Programme verfügen über eigene (zusätzliche) Formate
- Ist millionenfach genutzt → langfristige Unterstützung

**... Aber ...**

- Nicht jedes Programm kann alle Daten der ged-Datei verarbeiten
- Medien bedürfen einer speziellen Behandlung zur Einbindung



## GEDCOM – wie funktioniert es (I)

- GEDCOM Datei = Textdatei → mit Texteditor editierbar → \*.ged-Datei
- Daten sind gruppiert in Datensätze
  - HEAD           Kopf-Datensatz – 1x
  - INDI           Personen-Datensatz – 1/Person
  - FAM           Familien-Datensatz – 1/Familie
  - NOTE          Notizen-Datensatz \*
  - SOUR          Quellen-Datensatz \*
  - REPO          Aufbewahrungsort-Datensatz \*
  - OBJE          Medien-Datensatz \*
  - \_LOC          Lokation / Ort-Datensatz \*
  - TRLR          Abschluss – 1 Zeile
- Jeder Datensatz besteht aus mehreren Zeilen  
+ beginnt mit "0 @Xnn@ TTT"

```
0 HEAD
1 SOUR Ahnenforscher
2 VERS 2000
...
0 @I01@ INDI
1 NAME Karl Anton /Mustermann/
...
0 @F011@ FAM
1 HUSB @I01@
...
0 @N30@ NOTE Text
1 CONT Text
...
0 @S2@ SOUR
1 PAGE 4711
...
0 TRLR
```

\* nicht in allen Programmen enthalten



## GEDCOM – wie funktioniert es (2)

- Die wichtigsten Datensätze

- Personen-Datensatz

+

- Familien-Datensatz

```

0 @I01@ INDI
1 NAME Karl Anton /Mustermann/
1 SEX M
1 BIRT
2 DATE 15 DEC 1820
2 PLAC St. Gallen
1 OCCU Kaufmann
1 NOTE @N50@
  
```

```

0 @F011@ FAM
1 HUSB @I01@
1 WIFE @I13@
1 CHIL @I111@
1 MARR
2 DATE ABT 10 MAY 1855
2 PLAC St. Gallen
2 NOTE Bemerkung Ereignis
1 NOTE Bemerkung Familie
  
```

- Datensätze und deren Inhalt

- **Ziffern** geben die "Stufe" an, jede höhere Ziffer = Unterstufe der vorhergehenden
- **Tags** informieren über Art der nachfolgenden Information weitgehend standardisiert
- **Merkmale** enthalten die Information
  - Textphrase – z.B. Beruf, Ortsangaben, allgem. Texte
  - strukturiert entsprechend Vorgabe – z.B. Datum Ungenauigkeitsangaben
  - Zeiger zu anderen Datensätzen



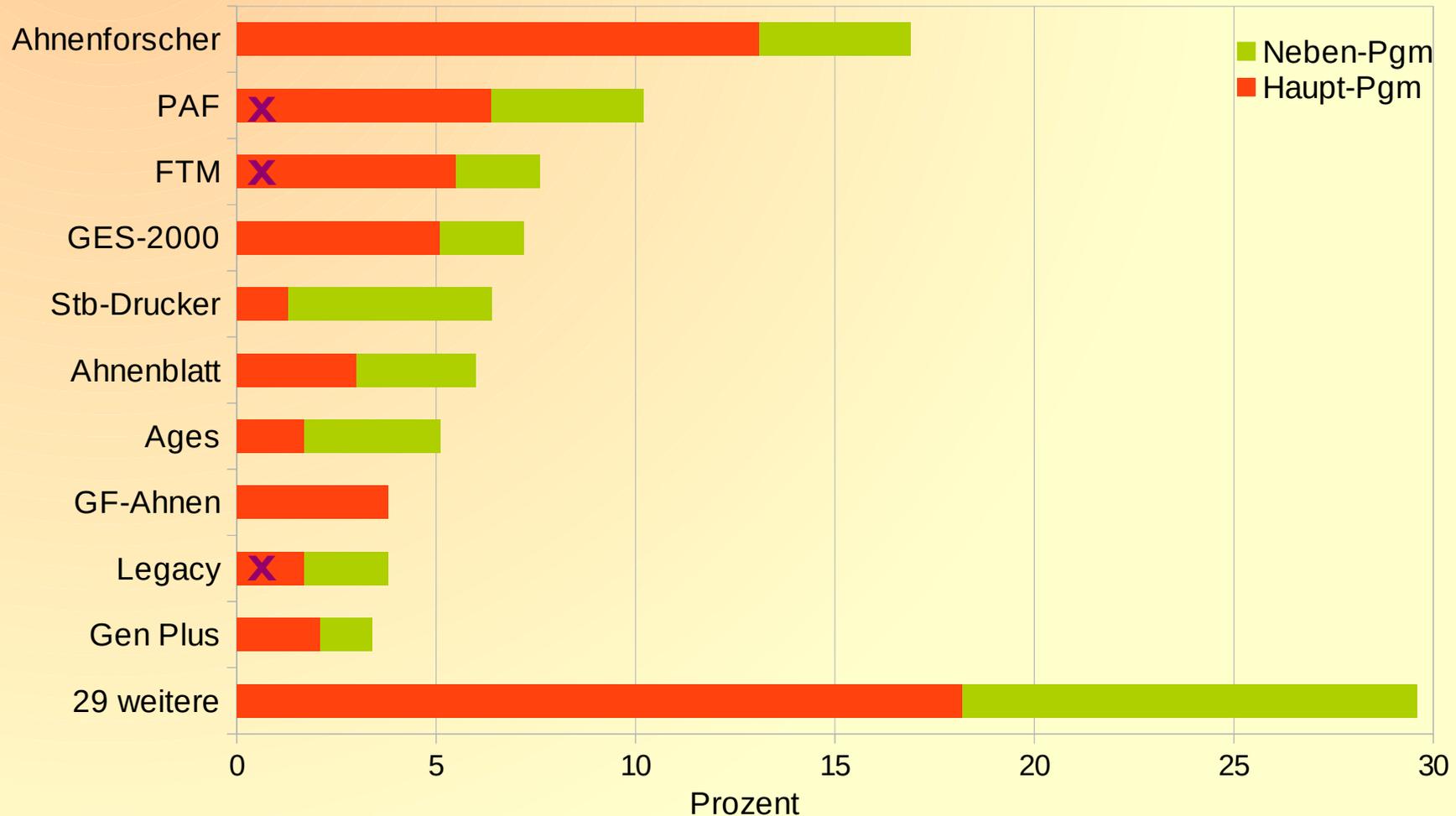
## Programme & GEDCOM

- Programme wurden zuerst erstellt
- GEDCOM folgte später
  - Zum Austausch von Daten
  - Basiert auf Forderungen von Programmen + Nutzer
  - Weiterentwickelt entspr. neuen Forderungen + technischen Möglichkeiten
- Heute gibt es einfache + komplexe Programme
  - Teilweise sind noch uralte Versionen in Gebrauch
  - Entsprechend bietet GEDCOM gewisse Flexibilität
- Programmentwickler legen fest
  - Datenfelder, deren Feldlänge und Zuordnung → Person, Ereignis, ...
  - Datenbank, Datenformate, Zeichensätze
  - Datenstrukturen → eingebettete Texte vs. eigenständige Datensätze
  - Eingabeprüfungen auf Zulässigkeit + logische Korrektheit
  - Export / Import in ged-Datei → was, wie, wo, welche Tags, ...



## Verwendete Programme Top-10 – SGFF Umfrage Hr. Widmer – Juni 2014

154 Antworten → 236 Nennungen



Alle in GEDCOM AG beteiligt ausser  
PAF, FTM, Legacy



## GEDCOM Flexibilität

- Standard Tags ( ca. 140 )
  - 40 Personen Ereignisse, 13 Familien Ereignisse, 7 Namen
- Nutzer definierte Tags "**\_XXX**" für zusätzliche Datenfelder
- Eingebettete Texte (einfach) vs. separate Datensätze (komplex)
  - Notizen (NOTE), Quellen (SOUR), Medien (OBJE), ...
- Mehrfach nutzbare Tags (Beispiele)
  - NAME für verschiedene Namen → Heirats-, Ruf-, Stamm-, Nickname, ...
  - BIRT für unterschiedliche Angaben in Quellen
  - MARR für standesamtliche und kirchliche Trauung
  - OCCU für unterschiedliche Berufe zu unterschiedlichen Zeiten + Orten
- Aber auch Defizite vorhanden
  - Reihenfolge von Ereignissen ohne Datum → Heiraten, Kinder, ...
  - Verknüpfung von Medien und deren Speicherung / Zuordnung
  - Gleichgeschlechtliche Partnerschaften, ...
  - Missbrauch von Datenfeldern durch Nutzer

Bsp:

Ahnenforscher: 5

Legacy: 53

BrothersKeeper: 47



## GEDCOM Flexibilität

### Datum

- **Standard**
  - n DATE Datum
  - Datum → "5 MAR 2000", "OCT 2000", "2000"
  - n DATE [FROM | TO | BEF | AFT | ABT | CAL | EST] Datum
  - n DATE FROM Datum TO Datum
  - n DATE BET Datum AND Datum
  - n DATE INT interpretiertes Datum (Freitext) blau = selten
  - n DATE (Freitext)
- Häufige Mängel bei der **Eingabe** sind Angaben wie
  - Ostern 2000, nicht verheiratet, ...
  - 10.\_\_.200x, 10. MAI 2000 ?, 1875/76, ...
- Häufige Mängel beim **Export** sind Angaben wie
  - 5.3.2000, 05.03.2000, ...
  - vor März 2000, zw. 15. und 20. Okt. 2000, ...



## GEDCOM Flexibilität

### Namen

- Standard

1 NAME Vornamen /Nachname/ angehängter Namensteil  
aber auch nur Teile wie "1 NAME Vornamen", "1 NAME /Nachname/", ...

2 GIVN Vornamen

2 SURN Nachname

2 TYPE [ aka | birth | immigrant | maiden | married | nutzerdef. ]

blau = selten

2 NICK Spitznamen

2 NPFX Dem Namen vorangestellte Namensteile (z.B. „Dr.“)

2 SPFX Dem Nachnamen vorangestellte Namensteile (z.B. „von“)

2 NSFX Dem Namen angehängte Namensteile (z.B. „jr.“)

2 NOTE Notizen Struktur

2 SOUR Quellen Zitierung

2 \_XXX Text

Anm: "BrothersKeeper" hat hier 19 nutzerdef. Tags

- Bei Mangel werden Nutzer kreativ und verwenden Kennzeichnungen

- " \* / - o. ähnlich am Anfang u/o Ende von Namensteilen für Spitznamen, Rufnamen, Namensänderungen, ...

- Was geschieht bei GEDCOM Export + Import ?



## GEDCOM Flexibilität

### Notizen

2 Optionen für Notizen (NOTE) → Eingebettet oder Datensatz

Notiz Strukturen: =

schwarz: beide

blau: nur Datensatz

#### Datensatz:

n NOTE @XREF:NOTE@  
+I SOUR Quellen Zitierungen



#### Datensatz NOTE:

0 @XREF:NOTE@ NOTE Text  
I CONC|CONT Text  
I REFN benutzerdefinierte Referenznummer  
2 TYPE benutzerdefinierter Referenztyp  
I RIN automatisierte Datensatz-ID  
I SOUR Quellen Zitierungen

#### Eingebetteter Text:

- ★ n NOTE Text
- ★ +I CONC|CONT Text
- +I SOUR Quellen Zitierungen

★ = Ahnenforscher → nur für Person + Familie

Komplexer bei Quellen



## GEDCOM Flexibilität

### Quellen

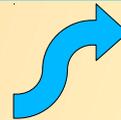
schwarz: beide

blau: nur Datensatz

Quellen Zitierungen: =

#### Datensatz:

```
n SOUR @XREF:SOUR@
+I QUAY [ 0 | 1 | 2 | 3 ]
+I OBJE <MULTIMEDIA_LINK>
+I NOTE Notiz Strukturen
+I PAGE Wo in der Quelle
+I EVEN [ADOP...WILL]
+2 ROLE [CHIL|HUSB|MOTH|(nutzerdef.)]
+I DATA
+2 DATE Aufzeichnungsdatum
+2 TEXT Textauszug der Quelle
```



#### Datensatz SOUR:

```
0 @XREF:SOUR@ SOUR
I TITL Quelltitel
I TEXT Text aus der Quelle
I DATA
2 EVEN Ereignistypen (ADOP...), kommasepar.
3 DATE [ FROM Datum ] [TO Datum ]
3 PLAC Ort
2 AGNC zuständiges Amt oder Institution
2 NOTE Notiz Strukturen
+ I AUTH Autor
I ABBR Quelle abgelegt unter / abgekürzt. Titel
+ I PUBL Publikationsdaten
+ I REPO @XREF:REPO@
2 CALN Ablagenummer, Bibliothekssignatur
3 MEDI [audio...video] Medientyp
2 NOTE Notiz Struktur
I REFN benutzerdefinierte Referenznummer
2 TYPE benutzerdefinierter Referenztyp
I RIN automatisierte Datensatz-ID
I NOTE Notiz Strukturen
```

#### Eingebetteter Text:

```
★ n SOUR Quellbeschreibung/-titel
+I TEXT Text der Quelle
+I QUAY [ 0 | 1 | 2 | 3 ]
+I OBJE <MULTIMEDIA_LINK>
+I NOTE Notiz Strukturen
```

★ = Ahnenforscher → für Person und \* ~ + = oo o|o



+ = Empfehlung GEDCOM AG: Übernahme in SOUR xxx

## GEDCOM Problematik

### Medien

- 2 Optionen für Medien (OBJE) → eingebettet oder Datensatz
  - Medien sind NICHT in der ged-Datei gespeichert, sondern nur deren Speicherort → "x FILE **Dateiname mit Pfad oder URL**"
  - Speicherort kann jedes Verzeichnis sein auf
    - Eigene Festplatte, externes Speichermedium, lokales Netz, Internet, ...
  - Entsprechend dem Speicherort ist die Referenz in der ged-Datei
    - Beachte:  
Die Dateistrukturen von Quell-System und Ziel-System sind normalerweise unterschiedlich
- ➔ Medien sollten mit der ged-Datei in das gleiche Verzeichnis oder (besser) einem direkten Unterverzeichnis kopiert und weiter gegeben werden



## GEDCOM Flexibilität

### Programme & Strukturen in ged-Dateien

"x" Datensatz "-" eingebettet

#	Programm	NOTE	SOUR	REPO	OBJE	_LOC
1	Ahnenforscher	-	-	-	-	-
2	PAF	x	x	x	-	-
3	FTM	x	x	x	x	-
4	GES	x	x	x	x	x
5	Stammbaumdrucker	-	-	-	-	-
6	Ahnenblatt	-	-	-	-	-
7	Ages	x	x	x	x	x
8	GF Ahnen	x	x	-	-	-
9	Legacy	x	x	x	-	-
10	Gen Plus	-	x	-	-	-

## GEDCOM Flexibilität

### Nutzer-definierte Tags

- Standard erlaubt dies für nicht abgedeckte Fälle
  - Tag muss mit "\_" beginnen → nicht jeder hält sich daran
  - Programme vergeben diese normalerweise  
Ausnahme: Reunion erlaubt Anwender diese zu vergeben → Wildwuchs
- Häufige Mängel
  - Programme liefern \_XXX Tags, obwohl Standard Lösung möglich
    - z.B. Ahnenforscher:
      - \_LEBENSORT Bern → RESI (mit TYPE Lebensort und PLAC Bern)
      - \_DIVERSES (mit TITL und NOTE) → EVEN/FACT (mit TYPE und NOTE)
  - Programme liefern z.T, keine Beschreibung der \_XXX Tags
    - ... verwenden zusätzlich Standard Sub-Tags mit anderer Bedeutung
    - ... verwenden selbst definierte Tag-Namen ohne "\_"



## GEDCOM Problematik

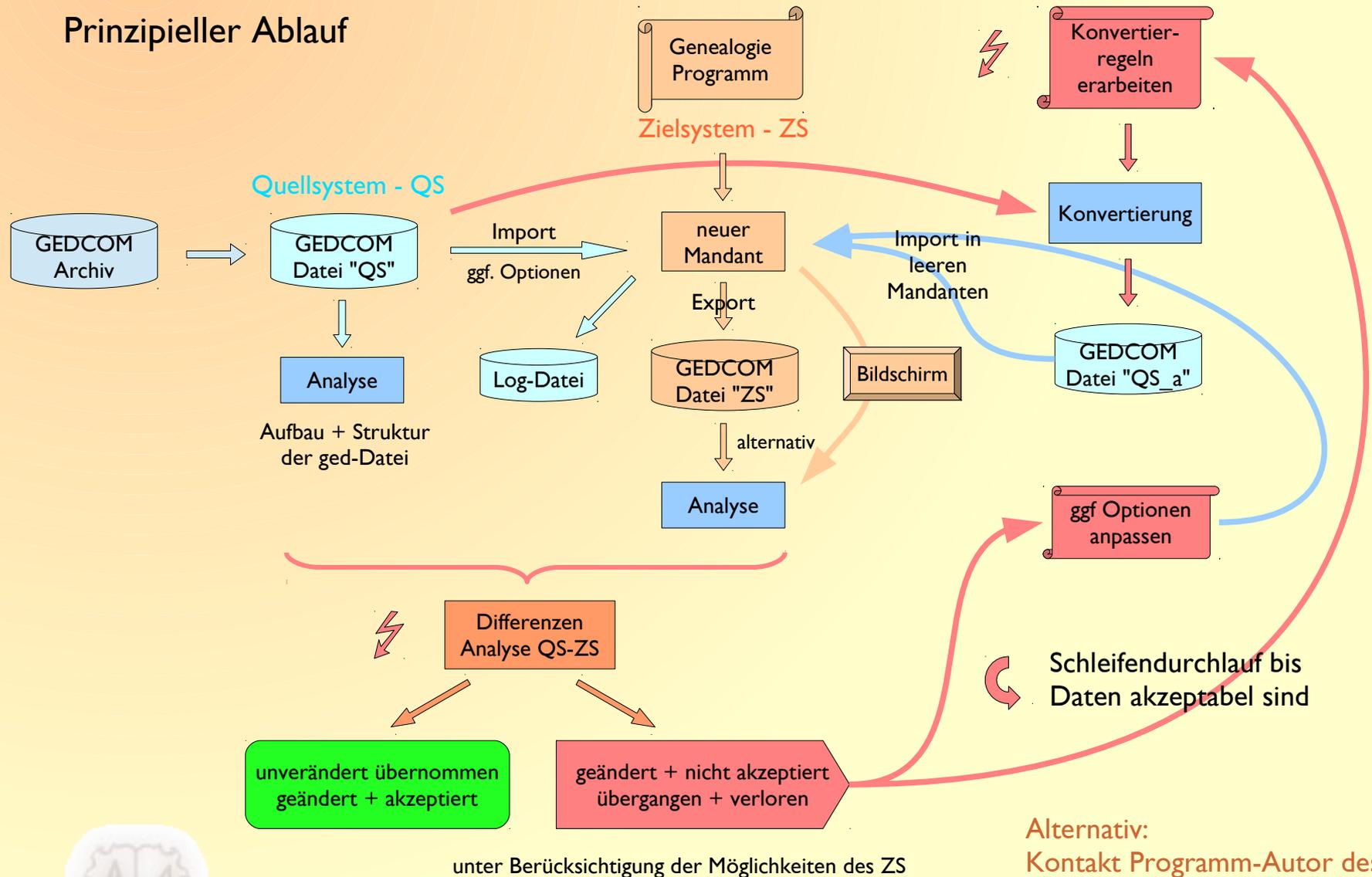
### Weitere Problemfelder

- Leerzeilen, eingerückte Zeilen
- Zeilenlängen > 255 Zeichen
- CONC Tag → eingeschobene Leerzeichen bei Verknüpfung der Zeilen
- Ungültige Zeichen → passen nicht zum Zeichensatz → betreffen meist Umlaute  
Verursacht durch "Copy & Paste" aus Dateien mit anderem Zeichensatz
- HEAD Datensatz → fehlende + falsche Angaben
- Stufe-0 Tags, aber nicht als Datensätze
- Steuerbefehle in Notizen → Tab, Zeilenvorschub, Zeilenende
- Formatierung, HTML-Befehle in Notizen → <b>, <i>, <pre>...
- Linkangaben in Notizen → http, href, ...
- Komplette GEDCOM Datensätze in Notizen



## Import von GEDCOM Dateien

Prinzipieller Ablauf



unter Berücksichtigung der Möglichkeiten des ZS

Alternativ:  
Kontakt Programm-Autor des ZS

## Welche Möglichkeiten bietet mein Programm

- Erstelle eine neue leere Genealogie-Datei
- Gebe min 5 Personen ein
  - 1 männlich (Ehemann) – mit allen möglichen Datenfeldern
  - 2 weiblich (Ehefrau + Liaison) – mit wichtigsten Datenfeldern
  - 2 beliebig (Kinder) – mit wichtigsten Datenfeldern
- Gebe 1 Heirat ein – mit allen möglichen Datenfeldern + 1 Kind
  - Gebe die Scheidung ein – mit allen möglichen Datenfeldern
- Gebe 1 uneheliche Verbindung ein – mit allen Datenfeldern + 1 Kind
  - Gebe die Trennung ein – mit allen möglichen Datenfeldern
- Eingabe / Verknüpfe Paten, Trauzeugen, Adoption, ...
  - Medien + andere Datenarten ggf nicht vergessen
- Export der Daten als ged-Datei + Analyse
- Optional kann auch eine existierende Genealogie so ergänzt werden, dass die oben angegebenen Kriterien bzgl. Vollständigkeit erfüllt sind



## GEDCOM Archivierung – was ist zu berücksichtigen (I)

### Fakten

- GEDCOM ist der einzige Standard → seit 30 Jahre  
Es gibt heute keine Alternative
- Es gibt unzählige ged-Dateien
- Jede Weiterentwicklung bzw. Neuentwicklung wird auf GEDCOM basieren → die interne Formatierung mag unterschiedlich sein
- Jede Änderung / Neuerung wird sich über Zeit etablieren oder in der Versenkung verschwinden
- Jeder Programm Autor wird bisherige und neue Formate über Jahre weiter pflegen (müssen)
- Auch neue Formate werden nicht ohne Probleme sein
- **Formate sind NICHT das Problem, sondern die verarbeitenden Programme und manchmal deren Nutzer**
- Formate sind unabhängig vom Speichermedium



## GEDCOM Archivierung – was ist zu berücksichtigen (2)

- Gute Dokumentation
  - Besonders für "Nutzer-definierte" Tags → gute Programme schreiben Erklärung in Kopf der ged-Datei.
  - Für missbrauchte Tags → viele Anwender schreiben alles mögliche in alle mögliche Felder und verwundern sich und andere.  
(Auch Programmierer missbrauchen Tags)
- Medien (von der ged-Datei referenziert)
  - Alle Medien (jpg, pdf, doc, xls, avi ... Dateien) werden in der ged-Datei zu ihrem Speicherort (Festplatte, Stick, Internet, ...) referenziert und sind selbst nicht in der ged-Datei enthalten.
  - Diese Medien sollten möglichst in einem Unterverzeichnis der ged-Datei enthalten sein und so referenziert werden.
- Speichermedium
  - Das gewählte Speichermedium sollte neben der ged-Datei die zugehörigen Medien und eine Beschreibung enthalten.



## GEDCOM Archivierung – was ist zu berücksichtigen (3)

- Empfehlungen
  - Jede Gruppe (Verein, Arbeitsgemeinschaft, ...) sollte möglichst 2 GEDCOM "Kenner" haben
  - GEDCOM Export möglichst als 5.5.1 (sonst 5.5) und UTF-8 Codierung
  - Bei Speicherung auf lokalen Datenträgern oder Server
    - Lagerung von Kopien an min. 2 verschiedene Orte
    - Alle 2-3 Jahre ein "Refresh" des Speichermediums
  - Oder Speicherung auf vertrauenswürdigen externen Server (Cloud)  
Ref: Vortrag "Umgang mit genealogischen Nachlässen" - Prof. Dr. Wulf von Restorff
- Weiterführende GEDCOM Literatur
  - GEDCOM 5.5 in englisch: <http://ofb.hesmer.name/files/gedcom/gedcom-55-en.pdf>
  - Deutsche Übersetzung der 5.5.1: <http://www.daubnet.com/ftp/gedcom-551-deutsch.pdf>
  - Deutsche Schnellreferenz: <http://www.daubnet.com/ftp/gedcom-schnellreferenz.pdf>
  - Detailbeschreibung der **GEDCOM AG** im GenWiki: <http://wiki-de.genealogy.net/Kategorie:GEDCOM-Tag>
  - Allgemeine Beschreibung der Tags in deutsch + englisch: <http://wiki-de.genealogy.net/GEDCOM-Tags>



**?? Fragen ??**

***Besten Dank für Ihr Interesse***

**Hinweise / Anregungen / Kritik / ...**

**bitte an**

**[ofb.hesmer@bluewin.ch](mailto:ofb.hesmer@bluewin.ch)**

**Link zum Vortrag:**

**[http://ofb.hesmer.name/files/pres/gedcom\\_bern\\_1410.pdf](http://ofb.hesmer.name/files/pres/gedcom_bern_1410.pdf)**

