

Suche in der Artland-Personendatenbank

Eine praktische Orientierungshilfe zur effektiven Ahnen- und
Regionalforschung

Status: Working Paper | Version 1.1

Publikationsdatum: Mai 2026

Autor: Ronald Voortman

Urheberrecht & Nutzungsbedingungen: Die in diesem Dokument veröffentlichten Forschungsergebnisse, Analysen und methodischen Aufarbeitungen unterliegen dem Urheberrecht von Ronald Voortman (Vortmes - Donkey Books). Die private Nutzung und das Zitieren im wissenschaftlichen Kontext unter Angabe der Quelle sind ausdrücklich erwünscht. Eine Vervielfältigung, Verbreitung oder öffentliche Wiedergabe zu kommerziellen Zwecken bedarf der ausdrücklichen schriftlichen Genehmigung des Autors. Die zugrunde liegenden Primärquellen sind gemäß den Richtlinien der jeweiligen Archivinstitutionen zu verwenden. Für die Korrektheit und Vollständigkeit der Daten wird keine Gewähr übernommen. Der

Suche in der Artland-Personendatenbank

Willkommen in der Personendatenbank des Artlandes auf **vortmes.nl**. Um Ihnen zu helfen, effektiv durch die Tausenden von historischen Personen, Rollen und Quellen zu navigieren, erklären wir im Folgenden die Funktionsweise und die Logik der Suchmaske

Einfache Suche versus Erweiterte Suche

Je nach Ausgangspunkt wählen Sie die Schnellsuche oder den gezielten Filter über mehrere Felder. Die Technik hinter diesen beiden Masken unterscheidet sich wesentlich, was direkten Einfluss auf Ihre Ergebnisse hat:

- **Einfache Suche (Das universelle Suchfeld):** Dieses System arbeitet mit einem einzigen Eingabefeld. Sie können hier mehrere Suchbegriffe eingeben (z. B. einen Vornamen und einen Nachnamen). Das System teilt Ihre Eingabe automatisch anhand von Leerzeichen auf und sucht in allen Indexfeldern über eine **UND-Verknüpfung** (**AND**).

Wie das funktioniert: Je mehr Suchbegriffe Sie eingeben, desto eingeschränkter und zielgerichteter wird das Ergebnis. Wenn Sie nach Jan Dallmann suchen, müssen sowohl der Begriff „Jan“ als auch „Dallmann“ irgendwo im Datensatz vorkommen. Ein „Jan Jansen“ oder ein „Berend Dallmann“ wird dadurch automatisch ausgeschlossen.

- **Erweiterte Suche (Gezieltes Filtern pro Feld):** Hier tragen Sie spezifische Begriffe in die dafür vorgesehenen Felder ein (wie „Familiennamen“ oder „Vorname“). Dieses System sucht standardmäßig nach einer exakten Übereinstimmung. Wenn Sie beim Familiennamen Dallmann eingeben, werden Personen mit der Schreibweise Dalman oder Dallman nicht angezeigt.

Wichtige Einschränkung: Keine Volltextsuche (No 'Full-Text Search')

Wichtige Einschränkung: Keine Volltextsuche (No 'Full-Tekst Search')

Die Datenbank unterstützt keine Volltextsuche. Dies bedeutet, dass ausschließlich in den spezifischen Indexfeldern der Suchmaske (wie Name, Rolle oder Kirchspiel)

Sie können daher **nicht** nach spezifischen Bauerschaftsnamen, Straßennamen, Berufen oder Todesursachen suchen, obwohl diese spezifischen Informationen zu einem späteren Zeitpunkt innerhalb des geöffneten Datensatzes durchaus sichtbar sein können.

Exakte Suche mit Wildcards versus Phonetische Suche

Wenn Sie die Erweiterte Suche wählen, ist die Suchmethode standardmäßig auf „**Exakte Suche (auch mit * und ?)**“ eingestellt. Da Namen in alten Registern oft unterschiedlich geschrieben wurden, können Sie diese exakte Suche durch den Einsatz von **Wildcards** (* und ?) flexibel gestalten.

⚠ Achtung: Das System schaltet nicht automatisch um. Um Wildcards zu verwenden, muss die Suchmethode explizit auf dieser exakten Einstellung stehen bleiben.

Das Sternchen (*) – Ersetzt null oder mehr beliebige Zeichen

Verwenden Sie *, wenn Sie den Anfang oder das Ende eines Namens wissen oder wenn mehrere Buchstaben in der Mitte variieren können.

Beispiele Für das Sternchen:

Suchanfrage Dallm* im Feld Familienname → Ergebnisse: *Dallmann, Dallmeyer, Dallmes.*

Suchanfrage *möl* im Feld Familienname → Ergebnisse: *Möller, Ostmöller, Mollermann.*

Das Fragezeichen (?) – Ersetzt genau ein Zeichen

Verwenden Sie ?, wenn Sie bezüglich eines bestimmten Buchstabens unsicher sind oder wenn ein Vokal wechselt. Dies ist ideal, um kleine Schreibweisenschwankungen abzufangen.

Beispiele Für das Fragezeichen:

Suchanfrage M?ler im Feld Familienname → Ergebnisse: *Moler, Müller* (aber nicht *Möller* oder *Mueller*, da dies zwei Zeichen sind).

Phonetische Suche (Phonetic Search): Als Alternative zur exakten Suche können Sie im Dropdown-Menü die phonetische Suche wählen. Das System sucht dann nach Namen, die wie Ihr Suchbegriff klingen, unabhängig von der Schreibweise. *Hinter den Kulissen wird derzeit an einer spezifischen Artland-Phonetik gearbeitet (abgestimmt auf den niederdeutschen/westfälischen Dialekt), zu der später mehr veröffentlicht wird.*

Zeitraum eingrenzen (Zeitraum)

Mit dem Zeitraum-Schieberegler über den Registern können Sie Ihre Suchanfrage direkt auf eine bestimmte historische Epoche beschränken (z. B. von 1591 bis 1857). Dies ist ideal, wenn Sie nach einem häufig vorkommenden Namen suchen, aber wissen, in welchem Jahrhundert sich Ihre spezifische Forschung abspielt. Datensätze außerhalb dieser Daten werden dann automatisch ausgeblendet.

Die Verwendung eines „Zweitpersonen-Vergleichs“

Manchmal suchen Sie nach einer Interaktion zwischen zwei Personen. Mit dem Zweitpersonen-Vergleich filtern Sie nach Dokumenten, in denen zwei bestimmte Personen zusammen in einem Ereignis (Event) genannt werden.

Wie das funktioniert: Dies ist besonders effektiv bei einer **Taufe** (Verbindung von Vater, Mutter, Kind oder Pate/Zeuge), einer **Heirat** (Braut und Bräutigam) oder bei einem **Todesfall** (der Verstorbene zusammen mit dem Ehepartner oder unter Nennung der Eltern). Dies gilt also nicht für allgemeine Angelegenheiten wie Erbfolgen.

Anwendungsbeispiel:

Sie suchen bei der ersten Person nach dem Familiennamen **Dallmann** und tragen bei der zweiten Person **Meier** ein. Sie erhalten dann direkt die Taufen, Heiraten oder Sterbeurkunden, in denen beide Namen zusammen eine Rolle spielen.

Rollen, Herkunftsorte und Register

In der erweiterten Suchmaske können Sie Ihre Suche basierend auf dem historischen Kontext verfeinern:

- **Rollen:** Dies ist die spezifische Rolle, die eine Person *zum Zeitpunkt der Eintragung* in der Quelle hatte (z. B. *Vater, Mutter, Kind, Braut, Bräutigam* oder *Zeuge*).
Achtung: Nicht jedes Register verwendet Rollen. Wenn eine Zelle in der Ergebnisliste leer ist, bedeutet dies lediglich, dass für diesen Dokumententyp keine spezifische Rolle erfasst wurde.
- **Herkunftsort:** Dieser Filter richtet sich nach den größeren **Parochien (Kirchspielen)** und ausdrücklich nicht nach den kleineren *Bauerschaften*. Dies hilft, die verschiedenen historischen Eingangsquellen klar zu identifizieren.

Beispiel: Jemand aus der Bauerschaft *Vehs* fällt unter das Kirchspiel *Badbergen*. Wenn eine Person aus *Vehs* (Badbergen) jemanden aus *Nortrup* (Kirchspiel *Ankum*) heiratet, wird für diese zweite Person Ankum als Herkunftsort registriert.

- **Registertypen auswählen:** Sie können ganz gezielt in bestimmten Registern suchen, indem Sie die Häkchen an- oder abwählen. Das System unterscheidet zwischen:
 - **Regionale Quellen:** Die lokalen Kirchenbücher (Taufregister, Heiratregister, Sterberegister).
 - **Externe Referenzen:** Register außerhalb der Region, in denen Artländer auftauchen (VOC Register, Poorter Register, Attest Register, Ondertrouw Register, Protocol Register, Auswärtiges Sterberegister).

Praxistipp:

Im Informationsfeld oben auf der Suchseite finden Sie einen direkten Link zu einer Übersichtsseite. Hier erhalten Sie detaillierte Einblicke in die Anzahl der Personen (Einträge) pro Herkunft pro Registertyp, welche Register bereits vollständig erfasst sind und welche sich derzeit in der Planung befinden.

Ergebnisse ansehen und tiefere Details aufrufen

Wenn Sie eine Liste mit Suchergebnissen erhalten, sehen Sie zunächst eine kurze Übersicht. Der eigentliche Informationsschatz verbirgt sich jedoch eine Ebene tiefer:

→ **Die Detailseite öffnen:** Klicken Sie in der Ergebnisliste ganz rechts auf den blauen Link des spezifischen Registers (z. B. *Heirat*, *Ondertrouw Register*, *Poorter Register*, *VOC Register*, *Auswärtiges Sterberegister* oder *Attest Register*).

Damit öffnen Sie den vollständigen Datensatz mit allen zusätzlichen Details, die in der Quelle gespeichert sind. Je nach Originaleintragung finden Sie hier die spezifischeren Daten, die nicht direkt durchsuchbar waren, wie **Alter**, **Todesursachen**, **Berufe** oder die genaue **Bauerschaft**. Bei bestimmten Registern (wie denen der VOC) enthält diese Detailmaske zudem einen direkten Hyperlink zum jeweiligen Archiv und – sofern verfügbar – einen Link zum **Originalscan** des Dokuments.

Artland-RMS

Artland-RMS steht für Artland Record Matching System. Es handelt sich um ein spezialisiertes Intelligenz- und Unterstützungsmodul innerhalb der WordPress-Archivumgebung.

Der Hauptzweck des RMS besteht darin, legitime Querverweise zwischen historischen Archivdatensätzen herzustellen, die sich auf dieselben Personen oder Familien beziehen. Da Namenseinträge in mittelalterlichen und frühneuzeitlichen Quellen stark variieren (durch phonetische Transkriptionen, dialektale Unterschiede und wechselnde Schreibweisen), wurde das RMS entwickelt, um auf Basis einer mehrkriteriellen, gewichteten Algorithmusanalyse automatisch eine Top 5 der wahrscheinlichsten „Schwester-Datensätze“ zu berechnen und dem Benutzer direkt zu präsentieren.

2. Funktionsweise und Architektur (Wie es funktioniert)

Das RMS wurde im Hinblick auf maximale Performance und eine schnelle Benutzererfahrung (User Experience) entwickelt:

- **Asynchrone Verarbeitung (Lazy Loading):** Beim Aufrufen einer Datensatzseite (sowohl über dynamische Detailseiten als auch über Custom Post Types/Pods) lädt zuerst die Quellseite. Das RMS führt die komplexe Matching-Berechnung im Hintergrund über eine AJAX-Anfrage aus. Dadurch erfährt der Besucher keine Verzögerung beim Laden der Hauptseite.
- **Zweistufiger Filteransatz:** Um die Datenbanklast bei einem Archiv von über 145.000 Datensätzen zu minimieren, arbeitet die Berechnung in zwei Phasen:
 - **SQL-Vorfilterung:** Zunächst wird eine Auswahl von maximal 200 potenziellen Kandidaten getroffen, die phonetisch über die Kölner Phonetik übereinstimmen.
 - **In-Memory-Scoring:** PHP berechnet anschließend über Levenshtein-Distanzen den exakten Matching-Score für diese Kandidaten.

3. Datennutzung: Datensatz vs. Index

Das RMS unterscheidet zwischen den Quelldaten des geöffneten Datensatzes und den Kandidatendaten im zentralen Index:

A. Daten aus dem geöffneten Datensatz (Quelle)

Wenn ein Besucher einen spezifischen Datensatz betrachtet, liest das RMS die folgenden Felder aus dem Quelldatensatz aus:

- **id / record_id / post_id:** Eindeutige Identifikation des Quelldatensatzes.
- **surname_kolner:** Der primäre phonetische Schlüssel nach der Kölner Phonetik für die SQL-Vorauswahl.
- **surname_meta, firstname_meta, patronym_meta:** Die normalisierten / Metaphone-Varianten der Namen.
- **origin_place_norm:** Der normalisierte Herkunftsort/Ortsname.
- **event_date:** Das Datum oder die Jahreszahl des registrierten Ereignisses.

B. Daten aus dem zentralen Index (wp_mig_index_data)

Der vollständige Suchraum besteht aus der indexierten Tabelle wp_mig_index_data. Aus den Kandidatendaten in diesem Index werden dieselben Felder abgerufen, um den Vergleich durchzuführen, ergänzt um Anzeigedaten für den Benutzer:

- **firstname_var, patronym_var, surname_var:** Die originalen/lokalen Namensvarianten für die Bildschirmanzeige.
- **registernaam & role:** Der Quellkontext und die Rolle der Person (z. B. Zeuge, Käufer, Verstorbener).

4. Berechnung des Matching-Scores

Der Matching-Score wird in Prozent (0 % bis 100 %) ausgedrückt und basiert auf einer gewichteten Summe von 5 Indikatoren. Jeder Indikator liefert einen Wert zwischen 0.0 (keine Übereinstimmung) und 1.0 (identisch).

Gewichtsverteilung (Gesamt = 1.00 / 100 %)

Feld / Faktor	Gewicht	Methode
Nachname (surname_meta)	40 % (0.40)	Levenshtein-Distanz auf phonetischem Meta-Schlüssel
Vorname (firstname_meta)	25 % (0.25)	Levenshtein-Distanz auf phonetischem Meta-Schlüssel
Patronym (patronym_meta)	15 % (0.15)	Levenshtein-Distanz auf phonetischem Meta-Schlüssel
Herkunftsort (origin_place_norm)	10 % (0.10)	Levenshtein-Distanz auf normalisiertem Ortsnamen
Zeitdifferenz (event_date)	10 % (0.10)	Linearer Verlauf auf Basis der Jahreszahldifferenz

Die mathematischen Formeln

1. Textuelle Übereinstimmung (Levenshtein auf META-Feldern)

Die textuelle Ähnlichkeit S_{Text} zwischen zwei META-Werten wird berechnet, indem die Levenshtein-Distanz L durch die Länge der längsten Zeichenkette M dividiert wird:

$$S_{\text{Text}} = 1 - \frac{L(\text{String}_1, \text{String}_2)}{\max(\text{Länge}(\text{String}_1), \text{Länge}(\text{String}_2))}$$

Beispiel: Ein Tippfehler oder eine kleine Variante im META-Code führt zu einem schrittweisen Absinken des Scores anstelle einer harten Ablehnung.

2. Übereinstimmung der Zeitdifferenz (S_{Jahr})

Historische Personen tauchen oft innerhalb eines bestimmten Zeitraums mehrfach auf. Die Jahreszahl-Übereinstimmung wird auf Basis der absoluten Differenz in Jahren ($\Delta y = |y_1 - y_2|$) berechnet:

$$S_{\text{Jahr}} = \begin{cases} 1.0 - (\Delta y \times 0.10) & \text{wenn } \Delta y < 10 \\ 0.0 & \text{wenn } \Delta y \geq 10 \end{cases}$$

(Falls bei einem der Datensätze kein Datum bekannt ist, erhält diese Komponente einen neutralen Score von 0.5).

3. Gesamter Matching-Score

Der definitive Matching-Score S_{Gesamt} ist die Summe der einzelnen Scores, multipliziert mit ihrer jeweiligen Gewichtung:

$$S_{\text{Gesamt}} = (S_{\text{Nachname}} \cdot 0.40) + (S_{\text{Vorname}} \cdot 0.25) + (S_{\text{Patronym}} \cdot 0.15) + (S_{\text{Herkunft}} \cdot 0.10) + (S_{\text{Jahr}} \cdot 0.10)$$

5. Schwellenwert und Ausgabe

- Schwellenwert (Threshold): Nur Ergebnisse mit einem $S_{\text{Gesamt}} \geq 0.50$ (mindestens 50 % Match) werden berücksichtigt.
- Sortierung & Präsentation: Die Kandidaten werden nach dem höchsten Score sortiert. Die Top 5 wird in der Benutzeroberfläche mit den dazugehörigen Matching-Prozentwerten und direkten Hyperlinks zu den betreffenden Detailseiten oder WordPress-Posts angezeigt.