



## **59. Deutscher Genealogentag**

14.–17. September 2007  
in Ludwigshafen/Rhein  
- Ortsbezüge in der Genealogie -

### **SlownikGeo**

Übersetzung eines 16-bändigen Ortslexikons  
(Polen und Osteuropa, 1880-1902)  
und Nutzbarmachung für genealogische  
Datenbanken

**Dr. Hanno V. J. Kolbe**

Hic Leones EURL, Achenheim, Alsace/France, Old Europe



## Übersicht

### **Słownik Geograficzny Królestwa Polskiego**

Am Anfang war das Buch ...

### **SlownikGeo**

Das Projekt, die Partner, das Know How ...

### **Vom Buch zur Datenbank**

Digitalisierung/Übersetzung von 14700 Seiten Text

### **Freier Flug der Fantasie ...**

An welches Buch denken Sie jetzt gerade ... ?

Die ersten drei Themen werden im Vortrag behandelt.

Mit dem letzten Punkt wird darauf hingewiesen, dass die logistischen und praktischen Werkzeuge, um ein solches Projekt durchzuführen, vom Verein für Computergenealogie e.V. und Hic Leones erarbeitet wurden und im Prinzip jedem Interessierten zur Verfügung stehen.



## **Słownik Geograficzny Królestwa Polskiego i Innich Krajów Słowiańskich**

### **Geographisches Lexikon des Königreiches Polen und anderer slawischer Länder**

16 Bände (14785 Seiten), Warszawa, 1880–1902

68 Hauptreferenzen (Bücher u. Zeitschriften)  
weitere Referenzen (ca. 500) im Text

677 Mitarbeiter unter Leitung von  
*F. Sulimerski*, B. Chlebowski u. W. Walewski

Wydawnictwa Artystyczne i Filmowe, Warszawa  
1986/87: 1000 Exemplare

S. Konarski (Genealoge): Namensindex  
Begonnen 1930's, beendet 1995 (S. Górzynski)

Um die Zahl von 14785 Seiten zu erläutern :

Würde man die Zeilen im Text aller 16 Bände in Originalgröße (etwa 10er Font) zu einem einzigen Textfaden hintereinanderfügen, erhielte man einen lesbaren Text von etwa 130 000 km Länge.

Dies entspricht einer Länge von mehr als 3 x um den Äquator ... und veranschaulicht die Menge an Information, die Sulimerski und Mitarbeiter in ihrem Werk verarbeitet haben.



**Im Słownik Geograficzny beschriebene Länder:**

- Polen**
- komplett
- Russland**
- Baltikum
- West-Gub.
- Süd-Gub.
- Deutschland**
- Westpreußen
- Ostpreußen
- Posen
- Pommern
- Schlesien
- Österreich**
- Galzien
- Schlesien
- Mähren
- Küstenland
- Ungarn (sl.)
- Bukovina (sl.)



**Im Słownik Geograficzny enthaltene Information:**

- Provinzen
- Kreise
- Städte
- Dörfer
- Weiler
- Höfe

**Polit./kirchl. Zuordnung**

**Demograph., relig. etc. Statistiken**

**Erziehung Industrie Landwirtsch. Handel Transport Geschichte 'Ereignisse' (Namen von Besitzern; seltener von Einwohnern)**

- Regionen
- Gebirge
- Flüsse
- Seen

Hier erfahren Sie, welche Bereiche das Lexikon abdeckt

- in der Hauptsache der hellgrüne Bereich -

und welche Informationen gefunden werden können.



## Partner im SlownikGeo-Projekt

Den Anfang  
machte die ...



Mehrere glückliche Ereignisse  
führten zu der Gestaltung des  
SlownikGeo-Projektes.



## Partner im SłownikGeo-Projekt

### **PGSA: Polish Genealogical Society of America**

(Polnisch-Genealogische Gesellschaft von Amerika)

Gegr. 1978

Etwa 2000 Mitglieder

Sitz in Chicago

Vier Publikationen pro Jahr

Jährliches Meeting in Schaumburg, IL

#### **Beitrag**

2003: Digitalisierung (Scans) der 16 Bände

Słownik Geograficzny Królestwa Polskiego

DjVu-Datei auf CD-ROM

Publication editor: William F. Hoffman

Editor: Rafał T. Prinke, Poznań



## SGKP DjVu-Version

**Słownik  
Geograficzny  
Królestwa  
Polskiego**

Select volume:  
1: [A - Der](#)  
2: [Der - G](#)  
3: [H - Kę](#)  
4: [Kęs - Kut](#)  
5: [Kut - Mal](#)  
6: [Mal - Net](#)  
7: [Net - Per](#)  
8: [Per - Poż](#)  
9: [Poż - Ruk](#)  
10: [Ruk - Soch](#)  
11: [Soch -  
Szu](#)  
12: [Szu - War](#)  
13: [War - Wor](#)  
14: [Wor - Żyż](#)  
15- [A - Jan](#)  
1: [A - Jan](#)  
15- [Jan - Ż](#)  
2: [Jan - Ż](#)

[Title page](#)  
[Documentation](#)  
[Plugin setup](#)

Copyright  
© PGSA 2003

**electronic facsimile edition**

---

**Słownik Geograficzny Królestwa Polskiego**  
i innych krajów słowiańskich

edited by  
**Filip Sulimierski, Bronisław Chlebowski and Władysław Walewski**  
Warszawa 1880-1902

---

conceived and edited by  
**Rafał T. Prinke**  
Poznań, Poland

with additional materials by  
**William F. Hoffman**

published by  
**Polish Genealogical Society of America**  
Chicago, USA  
2003

Dies zeigt die Eröffnungsseite der DjVu-Datei des Słownik Geograficzny.

Links kann der entsprechende Band ausgewählt werden.

Dann kann man in diesem digitalen Buch 'blättern'.

Die Datei enthält keine Suchfunktionen, da es sich um Bilddateien handelt.



## SGKP DjVu-Version

510

Pog

**Wyłkowyszek** 5 w., ma 3 dm., 28 mk. W 1827 r. było 3 dm., 15 mk.

**Pogramdyszki** al. *Pograndyszki*, pow. wyłkowyski, gm. i par. Wyłkowyszki, odl. od Wyłkowyszek 4 w., ma 2 dm., 29 mk. W 1827 r. było 3 dm., 15 mk.

**Pogramella**, zaśc., pow. wileński, w 3 okr. pol., odl. 5 w. od Wilna, 1 dm., 10 mk. katol.

**Pogramoncie**, ob. *Pogromoncie*.

**Pograndynie**, wś, pow. maryampolski, gm. i par. Wejwery, odl. od Maryampola 37 w., ma 9 dm., 70 mk.

**Pograndyszki** al. *Pograudyszki*, wś, pow. kalwaryjski, gm. Raudań, par. Kalwarya, odl. od Kalwaryi 10 w., ma 17 dm., 42 mk. W 1827 r. 5 dm., 46 mk. Wchodziły w skład dóbr rząd. i majoratu Kalwarya B.

**Pogranica**, zaśc. szl., pow. wileński, w 3 okr. pol., o 36 w. od Wilna, ma 1 dm., 9 mk. katol.

Pog

**ki**, gm. Baksza, par. prawosł. do wsi Płoski, o 29 w. od Bałty, ma 75 dusz męskich, 530 dzies. ziemi, osiedlona przez jednodworców; własność Sobańskich.

**Pogrejże**, wś i fol., pow. maryampolski, gm. Chlebiszki, par. Płutyszki, odl. od Maryampola 40 w., ma 2 dm., 16 mk. W 1827 r. należał do par. Preny, miał 1 dm., 19 mk.

**Pogrejże** al. *Juliszki*, dobra, pow. wilkomierski, par. Sudejki, własność Józefa Drozdowskiego.

**Pogređa**, wś rząd. nad pot. Skierz, pow. trocki, w 1 okr. pol., o 21 w. od Trok, 7 dm., 69 mk. katol.

**Pogrimmen** (niem ) 1.) dobra ryc. na prus. Litwie, pow. darkiejmski, o 7 klm. od st. p., tel. i kol. ż. Darkiejmy. Obszar 403 ha. 2.) **P.**, wś, tamże, st. p. Kleszczewo.

**Pogrobicie**, wś, pow. rossieński, par. betygolska.

Diese Abbildung zeigt das obere Drittel einer Seite.

(Sie können in den Text hinein-zoomen).

Man beachte die Scan-Qualität.

Die Abkürzung "dm." im Text bedeutet "dom" oder "domy" (poln. für "Haus" oder "Häuser").

Man findet dort Wohnplatzeinträge (z.B. der 2. von oben), die nur aus zwei Gebäuden (oder weiter unten) nur aus einem Gebäude bestehen.



## Partner im SłownikGeo-Projekt

- Scan: 16 Bände
- DjVu-Datei



... Potential des  
SGKP für deut.  
Ostforscher

Der zweite Projektpartner,  
Hic Leones, gesellt sich hinzu.



## Partner im SlownikGeo-Projekt

### *Hic Leones*

(Hier sind Löwen)

Gegr. 2003

Sitz in Achenheim/Elsass

Raison d'être: Geographische Dokumentation  
Historisch-Geographische Enzyklopädie, 3. Ed.  
Digitalisierung historischer Lexika  
Edition von PDF-Büchern

#### **Beitrag**

- Text-Digitalisierung aller Bände des  
Słownik Geograficzny Królestwa Polskiego
- Bereitstellung eines Editors
- 1000-seitiges PDF-Lexikon PL=>D (1872)



**Bild der DjVu-Datei**

=>

**Resultat nach OCR**

**Peterhof**, ob. *Piotrówka*, pow. bydgoski.  
**Peterhof** 1.) majątność tuż pod Tczewem, pow. starogardzki, obejmuje 193'77 ha obszaru. Osada ta należy pod względem administracyjnym do Tczewa, 2.) **P.**, wyb., pow. gdański, st. p. i kol. Pszczółki. 3.) **P.**, fol. do Orla należący, pow. grudziądzki. *Kt. Fr.*  
**Peterhof** (niem.), osada na pols. Warmii, pow. olsztyński, st. p., tel. i kol. żel. Olsztyn.  
**Peterhof**, pow. sycowski, ob. *Bielawki*.  
**Peterischken** (niem.), wś na prus. Litwie, pow. szyłokarczemski, st. poczt. Kallningken.  
**Peterkalk** al. *Budy*, wś, pow. wałecki, par. katol. Sypniewo, 8062'7 mr. obszaru. W 1868 r. 149 bud., 53 dm., 517 mk., 59 katol., 454 ewangelików.  
**Peterkaschuetz** (niem.), ob. *Piotrkosice*.  
**Peterkau**, niem., ob. *Piotrkowo, Pietrzykowy, Pietrzykówki*.  
**Peterkehnen** (niem.), wś na prus. Litwie, pow. wystrucki, blisko prawego brzegu rzeki Ditowy, st. poczt. i tel. Didlacken.

**Peterhof**, ob. *Piotrówka*, pow. bydgoski.  
**Peterhof** 1.) majątność tuż pod Tczewem, pow. starogardzki, obejmuje 193'77 ha obszaru. Osada ta należy pod względem administracyjnym do Tczewa, 2.) **P.**, wyb., pow. gdański, st. p. i kol. Pszczółki. 3.) **P.**, fol. do Orla należący, pow. grudziądzki. *Kt. Fr.*  
**Peterhof** (niem.), osada na pols. Warmii, pow. olsztyński, st. p., tel. i kol. żel. Olsztyn.  
**Peterhof**, pow. sycowski, ob. *Bielawki*.  
**Peterischken** (niem.), wś na prus. Litwie, pow. szyłokarczemski, st. poczt. Kallningken.  
**Peterkalk** al. *Budy*, wś, pow. wałecki, par. katol. Sypniewo, 8062'7 mr. obszaru. W 1868 r. 149 bud., 53 dm., 517 mk., 59 katol., 454 ewangelików.  
**PeterkHSchuetz** (niem.), ob. *Piotrkosice*.  
**Peterkau**, niem., ob. *Piotrkowo, Pietrzykowy, Pietrzykówki*.  
**Peterkehnen** (niem.), wś na prus. Litwie, pow. wystrucki, blisko prawego brzegu rzeki Bitowy, st. poczt. i tel. Didlacken.

Hic Leones digitalisiert den ersten Band des Lexikons und schickt den polnischen Text an die PGSA nach Chicago.

Daraufhin erteilt die PGSA Hic Leones das Recht, alle Bilddateien zu digitalisieren und erhält im Gegenzug von Hic Leones die gesamten 16 Bände im Textformat.

Durch diese Win / Win - Situation ist eine freundliche und sinnvolle Allianz entstanden.



## Partner im SlownikGeo-Projekt

- Scan: 16 Bände
- DjVu-Datei

Wegweisend in  
Genealogie,  
Informatik u.  
Kommunikation



- Digitalisierung
- Editor
- PL/D Lexikon

Als dritter Partner steigt der  
Verein für Computergenealogie e.V.  
in das Projekt ein.



## Partner im SlownikGeo-Projekt

# Verein für *Computergenealogie* e.V.

Gegr. 1989

Etwa 2800 Mitglieder

Sitz in Bremen

Mailing-Listen, GenWiki, GOV ...

Viermal/Jahr Computergenealogie

Ein jährliches Meeting

### **Beitrag**

- Import/Export Modul (Editor)
- Public Relations (Mailing-Listen)
- Informatik-Kompetenz



## Partner im SłownikGeo-Projekt

- Scan: 16 Bände
- DjVu-Datei



- Digitalisierung
- Editor
- PL⇒D Lexikon

- Editor Modul
- Public Relations
- Informatik



... bemüht um  
bessere  
Forschungs-  
möglichkeiten  
in Polen

Als vierter und letzter  
Partner kommt die  
Forschungsgruppe Grafschaft Glatz  
(FGG) hinzu.



## Partner im SlownikGeo-Projekt

### *FGG*

(Forschungsgruppe Grafschaft Glatz)

Gegr. 1986

FG in der AGoFF, 55 Mitglieder

Raison d'être: Kultur- u. Familiengeschichte in  
der Grafschaft Glatz

Kirchenbuchverkartung, Gedcom-Datenbank

Ein jährliches Meeting

### **Beitrag**

- Übersetzungsprogramm PL ⇔ D



## Partner im SłownikGeo-Projekt

- Scan: 16 Bände
- DjVu-Datei



- Digitalisierung
- Editor
- PL⇒D Lexikon

- Editor Modul
- Public Relations
- Informatik



- PL ⇔ D
- Übersetzungs-Software

Damit ist das Projekt-Team komplett und am

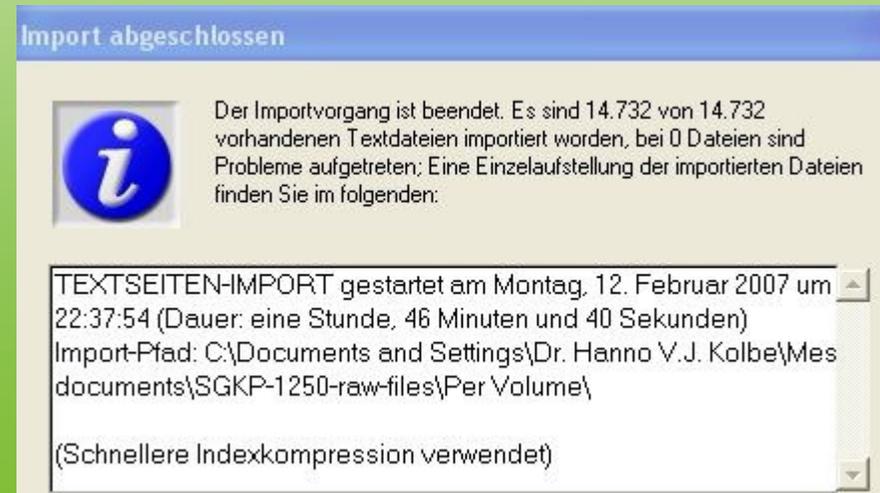
1. Jan. 2007

beginnt offiziell das Projekt SłownikGeo nach 18 Monaten Vorarbeit.



## Buch ⇒ Datenbank ⇒ Bearbeitung

- S/W-Scan jeder einzelnen Buchseite (⇒ Bild).
- Text-Erkennung (OCR-Software); Gruppen von etwa 200 - 400 Bildseiten.
- Abspeichern des erhaltenen Textes jeder einzelnen Buchseite in einem speziellen Format (1 Seite ⇒ 1 Datei).
- Automatisches Einlesen und Indexierung der einzelnen Textdateien in die Arbeitsdatenbank (d.h. Editor).
- Bearbeitung:
  - Formatierung
  - Korrektur
  - Übersetzung
  - Export/Import



Hier noch einmal in der Zusammenfassung der Ablauf des Buchseite-in-den-Computer-Vorganges.

Seit dem 12. Feb. 2007, 22:37h (s. Import-Screenshot), ist der komplette (polnische) Text des Słownik Geograficzny per Suchfunktion zugänglich.



## Bearbeitung: Planung

- **2007**
- Intern: Entfernung von OCR-Fehlern
- Intern: Formatierung
- Intern: Nummerierung
- Intern: Parsen von geograph. Begriffen
- Intern: Standardisierung geograph. Begriffe

malborski	Marienburg
malborskiego	Marienburg
malborskim	Marienburg
malborskich	Marienburg

- **2008**
- Kontakt mit der genealogischen Gemeinschaft
  - ohne PL-Kenntnisse (OCR-Korrektur)
  - mit PL-Kenntnissen (Übersetzung)

Zusätzliche Arbeit in der Standardisierung ergibt sich dadurch, dass jeder Name im Polnischen in sieben (7) Fällen dekliniert werden kann, was im Deutschen z.B. durch 'konstanter Ortsname + Präposition' erreicht wird.



## Inkrementelle (u. reversible) Übersetzung

**Mańkowo**, wś, pow. płocki, gm. Majki, par. Słupia, odl. o 28 w. od Płocka. Ma 15 dm., mk. 172, 630 mr. gruntu, 150 nieuż. W 1827 r. 14 dm., 67 mk.

**Mańkowo**, w|ś=>? **Df**, Kr. // BzHM. // Uj. // Pow. płocki, gm. Majki, par. Słupia, o|dl.=o=28=w.=o|d=>? **28 Werst (29,9 km) entf. von Płocka**. Ma 15 d|m.=>? **Gebäude**, mk. 172, 630 mr. gruntu, 150 nieuż. W=1827=r.=>? **Im Jahr 1827** 14 d|m.=>? **Gebäude**, 67 mk.

**Mańkowo**, **Df im Pow. Płock**, g|m.=>? **Gem.** Majki, p|ar.=>? **Kspl.** Słupia, o|dl.=o=28=w.=o|d=>? **28 Werst (29,9 km) entf. von Płocka**. Ma 15 d|m.=>? **Gebäude**, m|k.=>? **Ew.** 172, 630 mr. gruntu, 150 nieuż. W=1827=r.=>? **Im Jahr 1827** 14 d|m.=>? **Gebäude**, 67 m|k.=>? **Ew.**

**Mańkowo**, **Df im Pow. Płock**, **Gem. Majki**, **Kspl. Słupia**, **28 Werst (29,9 km) entf. von Płock**. M|a=>? **Er // Sie // Es hat 15 Gebäude**, 172 **Ew.**, 630 m|r.=>? **Morgen** g|runtu=>? **Boden // Land**, 150 n|ieuż.=>? **Brachland. Im Jahr 1827** 14 **Gebäude**, 67 **Ew.**

**Mańkowo**, **Df im Pow. Płock**, **Gem. Majki**, **Kspl. Słupia**, **28 Werst (29,9 km) entf. von Płock**, 15 **Gebäude**, 172 **Ew.**, 630 **Morgen Land**, 150 **Morgen Brachland; (1827)** 14 **Gebäude**, 67 **Ew.**

Hier wird durch fünf aufeinanderfolgende Schritte demonstriert, wie durch Wort-für-Wort Übersetzung (gefolgt durch Übersetzung längerer 'Textstrings') eine komplette Übersetzung eines Eintrags erreicht wird.

Diese Wort-für-Wort Übersetzung wird 'blind' in der gesamten Datenbank durchgeführt.

Der Ort Mankowo ist ein von etwa 360 000 Einträgen im Słownik Geograficzny.



# **Einige Blicke in die SlownikGeo Arbeitsdatenbank**

An dieser Stelle erhielten die Anwesenden die Möglichkeit

- (a) die aktuelle Strukturierung der Daten in der Datenbank zu sehen.
- (b) nach einem Familiennamen zu suchen.
- (c) nach einem Ortsnamen zu suchen.
- (d) nach allen Orten in einer Region zu suchen.



*Tut mir leid ... ich dachte, Sie wären schon fertig ...*

Leider ist die SlownikGeo-Datenbank noch nicht über das Internet einsehbar.

Aber das Projekt-Team arbeitet daran, die bereits jetzt komplett übersetzten Einträge (interessanterweise die kleinsten und oft schwer zu findenden Wohnplätze !) herauszuziehen und für das GOV und die Hic Leones Enzyklopädie aufzubereiten (geplant: Jan. 2008).



## Danksagung

**Cynthia Piech**, PGSA, Chicago, IL (*Kontakt*)

**Klaus-Peter Wessel**, CompGen (*Kontakt*)

**Wenzel/Welzel/Teuber**, Vorstand FGg (*Kontakt*)

**Peter Lingnau**, CompGen (*GOV-Partner*)

**Prof. Dr. Jürgen Udolph**, Univ. Leipzig (*Diskussionen*)

**Leon Stevens**, Cleveland, OH (*Glossar*)

**Jürgen Czerner** (*Poln. Text: SlownikGeo in GenWiki*)

Die hier zitierten Personen haben massgeblich zum Fortschritt des Projektes beigetragen.

Kontakt-Adresse für interessierte Mitarbeiter:  
kolbe@hicleones.com

Aktuelle Projekt-Beschreibung:  
<http://wiki.genealogy.net/wiki/SlownikGeo>