

The European Thesaurus on International Relations and Area Studies - A Multilingual Resource for Indexing, Retrieval, and Translation

Michael Kluck, Axel Huckstorf

Stiftung Wissenschaft und Politik (SWP) -
German Institute for International and Security Affairs
Ludwigkirchplatz 3-4, 10719 Berlin, Germany
michael.kluck@swp-berlin.org, axel.huckstorf@swp-berlin.org

Abstract

The multilingual European Thesaurus on International Relations and Area Studies (European Thesaurus) is a special subject thesaurus for the field of international affairs. It is intended for use in libraries and documentation centres of academic institutions and international organizations. The European Thesaurus was established in a collaborative project involving a number of leading European research institutes on international politics. It integrates the controlled terminologies of several existing thesauri. The European Thesaurus comprises about 8,200 terms and proper names from the 24 subject areas covered by the thesaurus. Because of its multilinguality, the European Thesaurus can not only be used for indexing, retrieval and terminological reference, but serves also as a translation tool for the languages represented. The establishment of cross-concordances to related thesauri extends the range of application of the European Thesaurus even further. They enable the treatment of semantic heterogeneity within subject gateways. The European Thesaurus is available both in a seven-lingual print-version as well as in an eight-lingual online-version. To reflect the changes in terminology the European Thesaurus is regularly being amended and modified. Further languages are going to be included.

1. Introduction

Although we live in an era of easy-to-use search engines and vast (but not comprehensive) online data, the need for controlled terminologies is increasingly voiced within the information community. Regardless the early sell-by date of some of this information, it is still and with renewed emphasis necessary to get all the information overload and floods of data structured and organized. Thus, thesauri and other forms of documentary languages are even in the age of internet an important means for describing, storing and retrieving of information. And, apart from that, they can furthermore be used as terminological reference works for a specific subject area, or, in the case of a multilingual vocabulary, as tools for supporting translation work. They also can be used as the fundamental tool for the treatment of semantic heterogeneity. Finally multilingual thesauri are integrating different cultural and political views into one conceptual framework (Hudon, 1997).

The *European Thesaurus on International Relations and Area Studies* (European Thesaurus) covers a specific subject field within the social sciences, and more specific within the political sciences, that becomes increasingly important for the political life of countries and people. It has been developed within a truly European cooperation, in the sense of a quite complicated, long lasting and strongly discursive process. But in the end it was an integrative and successful process of cooperation where the integration of languages with all the peculiarities and political cultures of different countries were finally managed (Seydel, 2004).

2. The Development of the European Thesaurus

The multilingual special subject *European Thesaurus on International Relations and Area Studies* has been developed in the framework of the *European Information Network on International Relations and Area Studies (EINIRAS)*, a cooperation project set up in the 1990s between the library and information departments of specialist European research institutes in these fields.¹ It is part of an effort to explore and devise methods and systems for the participating institutions to work more closely together in providing documentation and information services on all aspects of international and foreign affairs like

- international relations / foreign policy of individual states
- international security / military affairs / security policies of individual states
- international cooperation and international organizations
- development policy and development policy cooperation

¹ The following EINIRAS member institutes participated in the European Thesaurus project: Deutsch-Französisches Institut, Ludwigsburg; Otto Suhr Institute for Political Science, Freie Universität Berlin; Fundació CIDOB, Barcelona; GIGA German Institute of Global and Area Studies, Hamburg; Hellenic Foundation of European and Foreign Policy, Athens; Institute of Scientific Information in the Social Sciences, Moscow; Istituto Affari Internazionali, Rome; Polish Institute of International Affairs, Warsaw; Royal Institute of International Affairs, London; Stiftung Wissenschaft und Politik, Berlin; Institute of International Relations, Prague.

- international economics
- European integration politics
- transatlantic relations
- regional and country studies worldwide
- foreign cultural policy
- international law

2.1. Source Thesauri

The European Thesaurus integrates the concepts and terminologies of a number of specialist thesauri currently used in research and information units across Europe. The main sources were:

- CERES Thesaurus - Council of Europe
- The Royal Institute of International Affairs Library Thesaurus
- Thesaurus for Stockholm International Peace Institute Library Catalogue and Article Database
- Thesaurus de l'IFRI (Institut français des relations internationales)
- Thesaurus of the German Information Network International Relations and Area Studies (Fachinformationsverbund Internationale Beziehungen und Länderkunde, FIV)

For the field of international law the subject heading lists of the library of the Max Planck Institute for Comparative Public Law and International Law in Heidelberg, Germany, and of the Peace Palace Library in The Hague, Netherlands, were consulted.

2.2. The Path to the European Thesaurus

These various related collections of international affairs concepts were used in a multi-stage process of long duration to identify the relevant terminology for the new integrated thesaurus on international relations and area studies and to define the meaning of descriptors. Besides examining the mere terminologies, the contents of the databases indexed by these different thesauri or subject heading lists were taken into consideration for the decision on relevant descriptors. Even searches with search engines were carried out to get an estimate of the usage of terms and term variants under discussion.

As a first work step the quite diverse subject 'International Relations and Area Studies' had to be subdivided into fields of smaller scope (e.g. International Security/Military realm, International Law etc.). This was done mainly following the existing thematic structure of the thesaurus of FIV. After this necessary differentiation had been done the descriptors chosen for the European Thesaurus were assigned to either one or more of these sub-fields. In most cases several denominations existed for one concept so that a decision had to be taken on which denomination should become the preferred term (descriptor) and which denomination should stay as non-preferred term (non-descriptor). An important crite-

riterion was the commonness of a term both in academic writing as well as in political practice.

The English denomination of a descriptor was usually the basis from which the equivalents in the other languages had been determined. This work was done by native speakers from the participating institutes. In a few cases, mostly concerning the Eastern European languages, equivalent terms did not (yet) exist in the respective languages. Thus, besides translations according to existing dictionaries 'artificial denominations' had to be created.

To enhance the access to the controlled vocabulary of the European Thesaurus numerous non-descriptors were included. Due to the specifics of the languages represented this entry vocabulary can vary for each language.

The project's work was supported by a terminology database accessible for all project members. There was one dataset for each descriptor containing all relevant information such as: equivalents in the other languages (as English technically functioned as the preferred term), non-preferred terms (synonyms and abbreviated forms of terms), scope notes, assignment to thematic fields, notations, cross-concordances with source thesauri, administrative information.

2.3. Updating the European Thesaurus

Concepts and terms are subject to permanent, but moderate change. The *European Thesaurus Development Group*, which is composed of representatives from those institutes using the European Thesaurus for indexing and retrieval in their respective databases (see chapter 4.3), constantly monitors the developments of the terminology in the area of international relations and modifies and amends the European Thesaurus regularly with about 50 changes and new entries every year.

2.4. Cross-concordances with other Controlled Terminologies

In addition cross-concordances to other existing thesauri have been intellectually constructed after the completion of the European Thesaurus. At the moment these are the *PAIS Subject Terms Thesaurus* (English), the *Thesaurus for Economic and Social Development* (TWSE, German), the *Standard Thesaurus Economics* (STW, German, English), the *Thesaurus for Social Sciences* (Thesaurus Sozialwissenschaften, German, English, Russian). These cross-concordances will be used to support the simultaneous search of heterogeneously indexed bibliographical data in multiple databases. The mappings allow the user to search with terms from one terminology getting results also from databases using another of these mapped terminologies (Kluck, 2007; Mayr, 2006). Cross-concordances can thus be seen as a means to "translate" one controlled terminology into another.

3. The Structure of the Multilingual European Thesaurus

English was selected as the primary or source language. Equivalents of the English descriptors are available in the six target languages (print version): Czech, French, German, Italian, Polish and Spanish. The online version already contains an additional language: Croatian. As the ninth language Russian will be added in the beginning of 2008.

The European Thesaurus contains approximately 8,200 descriptors (preferred terms) and numerous non-descriptors (synonyms or abbreviated forms). It provides a well-structured and wide-ranging "terminological tool" of controlled vocabulary to describe the subject content and format of all kinds of literature and documents (books, reports, essays, periodicals, newspaper-articles, texts of agreements etc.) in databases dealing with international politics, international security, international economics and area studies.

As already mentioned, the terminology of the European Thesaurus is assigned to 24 micro-thesauri or subject fields. Within those micro-thesauri a further differentiation has been made. In so-called field-groups all related terms of a sub-theme of a field are subsumed. However, a associative relationship in the strict sense does not necessarily exist between the descriptors within a sub-field.

Due to the nature of social sciences terminology as well as to limit the number of descriptors the European Thesaurus contains a great amount of compound terms, especially of pre-coordinated terms.

Most of the descriptors deal with concepts of interest to researchers and practitioners in the field of international relations, international security issues and area studies (ideas, processes, agents, institutions, objects etc.). The European Thesaurus includes not only subject terms but also a number of proper names and geographical denominations to enable better-focused subject content description, for instance geographical place names, titles of international agreements/treaties, names of institutions / organizations.

4. Use Cases and Accessibility of the European Thesaurus

The European Thesaurus is currently available in a printed version (covering seven languages) and an online version (covering eight languages). In 2008 the Russian language will be added to the online version.

4.1. Print Version of the European Thesaurus

The printed version of the European Thesaurus consists of:

- a) a systematic volume in which the descriptors, in all seven languages in parallel, are set out in logical groups and sequences within a series of broad subject fields or "chapters",
- b) seven single-language volumes, each listing in alphabetical order the descriptors and selected non-descriptors in one language, indicating their position in the systematic structure.

The print version gives less information than is to be found in the online version of the thesaurus; in particular it includes only a selection of non-descriptors and it does not show the hierarchical and associative relationships between the terms. Here, the systematisation of the descriptors has been implemented through the aforementioned structure of subject fields and further sub-groups within the subject fields (Seydel et al., 2006).

4.2. Online Access to the European Thesaurus

The European Thesaurus (including full semantic relationships) is publicly available on the internet via the *International Relations and Area Studies Gateway* (www.ireon-gateway.eu). This new internet portal on international relations was launched in November 2007, and is also part of the German network of scholarly internet portals (www.vascoda.de). The online version of the European Thesaurus enables the user to search for relevant descriptors or non-descriptors by using any of the eight languages for this search. The descriptor or non-descriptor found will be displayed with all its semantic relations, scope notes (if available) and the subject field-group(s) to which the term has been assigned to. If needed the translations of the descriptors (including all relationships) will be shown by selecting the respective language(s). On the right hand side of the screen the user can browse alphabetically or systematically (according to the subject fields) through the thesaurus (see Figure 1 showing the example of "foreign policy" at the keywords page of the International Relations and Area Studies Gateway, beta version).

At the *Gateway International Relations and Area Studies* the online version of the European Thesaurus can be used twofold. First you can search within the thesaurus itself: to search for relevant descriptors, to look up a descriptor's semantic relationships or its translations. Second the search within the gateway of International Relation and Area Studies can be supported by the European Thesaurus: by searching and marking thesaurus term you can initialize a search within the gateway itself. Thesaurus terms can be searched under "Keywords" within the European Thesaurus and then be shifted to the search field (directly starting a search) by clicking a search start button. Otherwise you can take a thesaurus term from a list of keywords that also includes also non-descriptors and other additional controlled terms, and then search the World Affairs Online database (WAO).

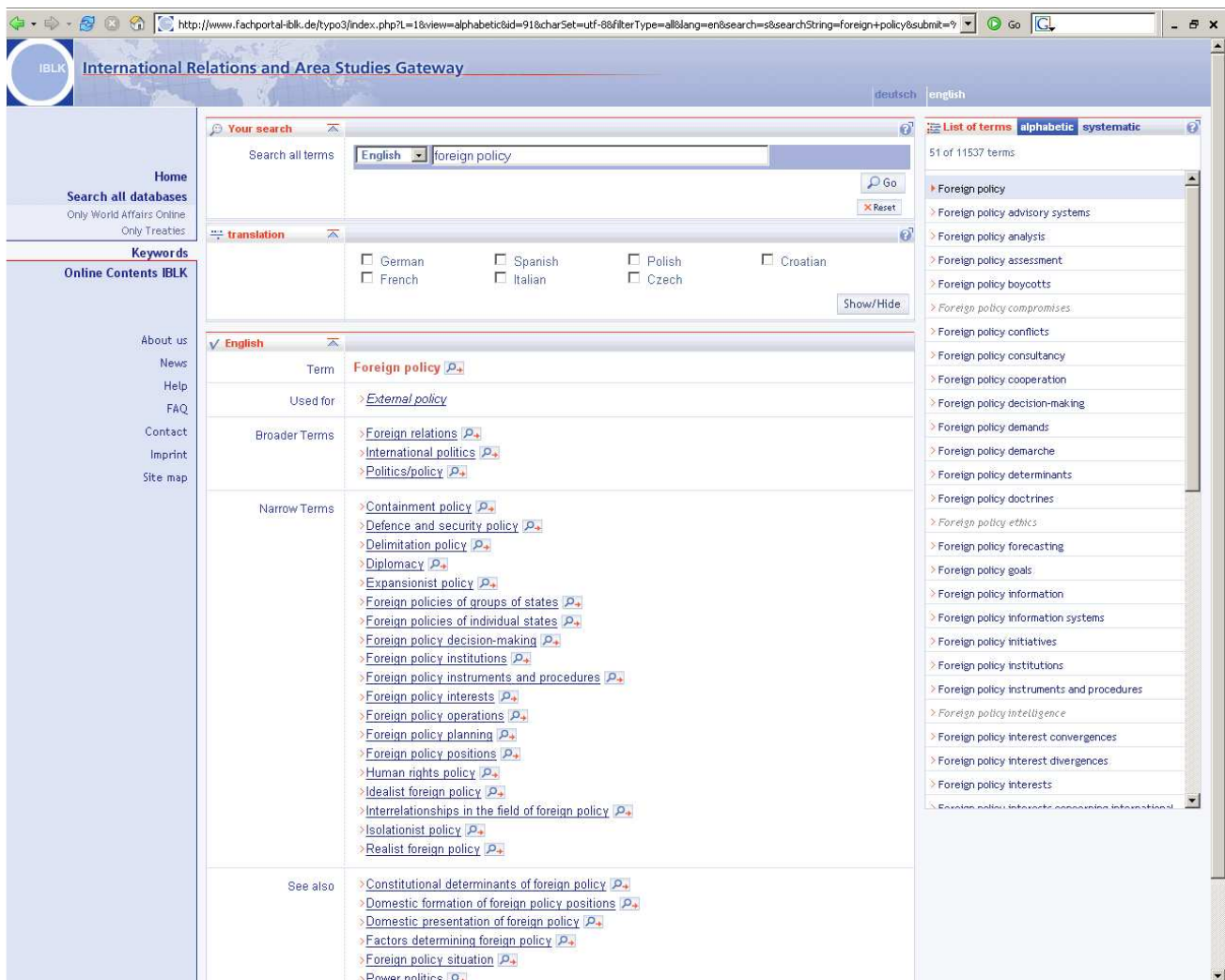


Figure 1: Example of descriptor display (online version of the European Thesaurus at the International Relations and Area Studies Gateway)

4. 3 The European Thesaurus as Indexing Tool

The European thesaurus is currently used as indexing tool for the production of several databases. The German Information Network on International Relations and Area Studies (FIV = *Fachinformationsverbund Internationale Beziehungen und Länderkunde*) is using the thesaurus for in-depth indexing of the database *World Affairs Online* (WAO) that can be accessed via the *International Relations and Area Studies Gateway* on international relations and area studies (www.ireon-gateway.eu). Within the German Information Network (FIV) thirteen institutes are collaboratively producing the database WAO. Here the mainly used indexing language is German. Nevertheless the users can search the database WAO in any of the thesaurus languages

The *Polish Institute of International Affairs* (PISM in Warsaw) is indexing its database with the thesaurus using mainly the Polish terms of the thesaurus. The *Czech Institute on International Relations* (IIR in Prague) uses mainly the Czech descriptors for indexing their database.

4. 4 Solving the Problem of Semantic Heterogeneity

The European Thesaurus and the cross-concordances to other thesauri are the basic elements to solve the problem of semantic heterogeneity (Walter et. al. 2006). The translations of terms (descriptors and non-descriptors) within the multilingual thesaurus as well as the cross-concordances to other thesauri are prerequisites to set up a heterogeneity service (HT service): the HT service transforms a given query into all equivalents of the contained terms within the given thesaurus that is attached to a certain database and within the corresponding thesauri of other databases. In such way a query A expressed in the query vocabulary A is directed to the subject gateway, sent to the HT service, and transformed into the entry vocabularies A, B, C of all connected databases A, B, C. In the example of Figure 2 the entry vocabulary A equals the query vocabulary A (and only needs the expansion to translations and equivalent terms), whereas for the databases B and C the query has to be transformed into the different index

ing languages or entry vocabularies B and C (see figure 2, Mayr/Walter, 2007; Kluck 2007).

The heterogeneity service² is a pre-processing step before starting the search engine. The terms of the original query are augmented by the transformations into other vocabularies/thesauri and by the expansions into all translations and into all associated non-descriptors. Thus, a given query term can not only be used for searching within the different descriptor fields of the documents, but also for searching within the whole amount of textual information being converted into the basic index of each of the queried databases (Mayr et al. 2008).

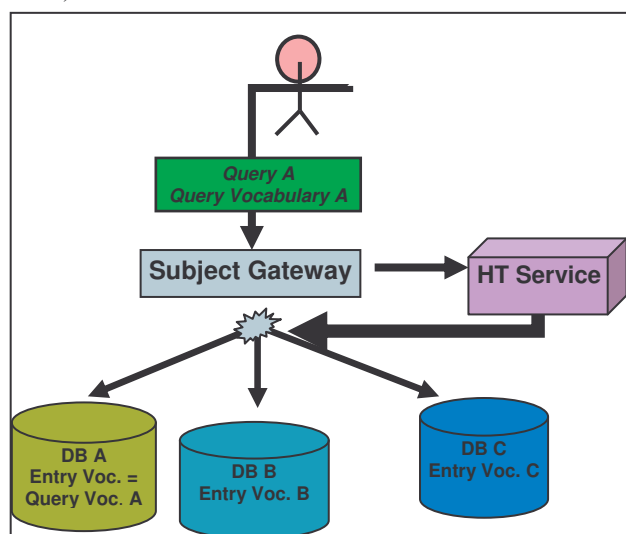


Figure 2: Workflow of the heterogeneity service

4.5 The Multilingual European Thesaurus as Translation Tool

Furthermore the multilingual European thesaurus may be used as a translation tool, because it comprises a lot of specific scholarly vocabulary and its mapping into other languages. This means not only translation of “words”, but also transformation of concepts into other cultural and political contexts.

5. Conclusion

The European Thesaurus on International relations and Area Studies demonstrates that European collaboration efforts result in a valuable tool not only for database production, but also for supporting search activities and research on the web in an intelligent way. This thesaurus provides not only a multilingual translation vocabulary, but also a systematic multilingual view of a certain discipline.

² The heterogeneity service was developed by GESIS – IZ (GESIS - Social Science Information Centre), Bonn, Germany within the project Competence Center Modeling and Treatment of Semantic Heterogeneity. GESIS-IZ is also providing the technical platform for the International Relations and Area Studies Gateway.

In the age of the World Wide Web such an intellectual work and costly project provides a powerful tool for reducing information overload and targeting on scholarly information. By connecting adjacent disciplines that cultivate different types of elocution and wording through an heterogeneity service users are fairly supported in digging into different areas of scholarly information. But this service needs the backing of well developed concept spaces that are represented by the thesauri in use and further development and their mapping into each other.

6. Acknowledgments

First of all we thank all partners of the European Network on International Relations and Area Studies for their intense work on the European Thesaurus. Particularly the following colleagues had been involved in the thesaurus development : Dietrich Seydel (SWP, Berlin), Susan J. Boyde (RIIA, London), Leszek Cyrzyk (PISM, Warsaw), Matthias Bauermeister (SWP), Alessandra Bertino (IAI, Rome), Laia Bonet I Bagant (CIDOB, Barcelona), Mariza Cricorian (IAI), Gerda Hansen (GIGA, Hamburg), Volker Hornung (FUB, Berlin), Helena Kolátorová (IIR, Prague), Bérénice Manac’h (DFI, Ludwigsburg), Dieter Menyesch (DFI), Gottfried Reinknecht (GIGA), Malgorzata Timoszuk (PISM).

Further on we have to thank Maximilian Stempfhuber and Philipp Mayr (GESIS-IZ, Bonn) for the development and implementation of the heterogeneity service as part of the International Relations and Area Studies Gateway which is also technically supported and hosted by GESIS-IZ. By running the International Relations and Area Studies Gateway we can heavily re-use a lot of research and software development done by GESIS-IZ

7. References

- Hudon, Michéle (1997). Multilingual thesaurus construction - integrating the views of different cultures in one gateway to knowledge and concepts. in: *Information Services & Use*, 17 (1997), 2/3, pp. 111--123.
- Kluck, Michael (2007). Die semantische Heterogenitätsbehandlung im Fachportal Internationale Beziehungen und Länderkunde (IBLK). In: Ockenfeld, Marlies (Ed.): *Information in Wissenschaft, Bildung und Wirtschaft. 29. Online-Tagung der DGI, 59. Jahrestagung der DGI, Frankfurt am Main 10. bis 12. Oktober 2007, Proceedings*. Frankfurt/Main: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis, pp. 247--254.
- Mayr, Philipp (2006). Thesauri, Klassifikationen & Co - die Renaissance der kontrollierten Vokabulare? In: Hauke, Petra; Umlauf, Konrad (Eds.), *Vom Wandel der Wissensorganisation im Informationszeitalter. Festschrift für Walther Umstätter zum 65. Geburtstag*. Bad Honnef: Bock + Herchen Verlag, pp. 151--170.

<http://edoc.hu-berlin.de/miscellanies/vom-27533/151/PDF/151.pdf>

Mayr, Philipp; Mutschke, Peter; Petras, Vivien (2008 to appear): Reducing semantic complexity in distributed Digital Libraries: treatment of term vagueness and document re-ranking. In: *Library Review*, 57 (2008) 3.

Mayr, Philipp; Walter, Anne-Kathrin (2007): Einsatzmöglichkeiten von Crosskonkordanzen. In: Stempfhuber, Maximilian (ed.): *Lokal - Global: Vernetzung wissenschaftlicher Infrastrukturen*; 12. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland [2006]. Bonn: GESIS - IZ Sozialwissenschaften. S. 149-166.

<http://www.gesis.org/Information/Forschungsuebersichten/Tagungsberichte/Vernetzung/Mayr-Walter.pdf>

Seydel, Dietrich (2004). Mehrsprachiger „European Thesaurus International Relations and Area Studies“ als Basis für europäische Zusammenarbeit zwischen Fachbibliotheken. Ein Werkstattbericht. In: *Das Kompetenzzentrum, Festschrift Udo Rossbach, 21. Oktober 2004* / Institut für Auslandsbeziehungen, Stuttgart pp. 31--46.

<http://www.ifa.de/w/download/festschrift.pdf>

Seydel, Dietrich; Boyde, Susan J.; Cyrzyk, Leszek (Eds.) (2006): *European Thesaurus on International Relations and Area Studies / Thésaurus Européen Relations internationales et études régionales / Europäischer Thesaurus Internationale Beziehungen und Länderkunde / Tesauro Europeo Relaciones Internacionales y Estudios Regionales / Tesauro Europeo di Relazioni internazionali e studi regionali / Evropský tezaurus mezinárodních vztahů a regionálních studií / Europejski Tezaurus Stosunków Międzynarodowych i Studiów Regionalnych*, Multilingual systematic part and monolingual alphabetical parts, 3 vol., Berlin: Stiftung Wissenschaft und Politik.

Walter, Anne-Kathrin; Mayr, Philipp; Stempfhuber Maximilian; Ballay, Arne (2006). Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung in Informationssystemen. In: Stempfhuber, Maximilian (Ed.): *In die Zukunft publizieren: Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften*; 11. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland. Bonn: IZ Sozialwissenschaften, pp 205 -- 226.

http://www.gesis.org/information/forschungsuebersichten/tagungsberichte/publizieren/IuK_Tagungsband_11_Walter.pdf

8. Links

Fachinformationsverbund Internationale Beziehungen und Länderkunde (FIV) (German Information Network International Relations and Area Studies (FIV), coordinated by Stiftung Wissenschaft und Politik (SWP), Berlin, Germany
www.fiv-iblk.de

Institute of International Relations (IIR) Prague, Czech Republic
www.iir.cz

Polish Institute of International Affairs (PISM) Warsaw, Poland
www.pism.pl

Stiftung Wissenschaft und Politik (SWP) Berlin, Germany
www.swp-berlin.org

IREON Gateway International Relations and Area Studies Online (in German: **IREON Portal** Internationale Beziehungen und Länderkunde Online), published by SWP, Berlin, Germany³
www.ireon-gateway.eu or www.iras-portal.eu

European Thesaurus at the IREON Gateway
www.iras-portal.eu/index.php?id=91&L=1

EINIRAS – European Information Network on International Relations and Area Studies
www.einiras.org

German Scholarly Information Network **VASCODA**
vascoda e.V., Hanover, Germany
www.vascoda.de

GESIS - Social Science Information Centre (GESIS-IZ), Competence Center Modeling and Treatment of Semantic Heterogeneity, Bonn, Germany
http://www.gesis.org/en/research/information_technology/komohe.htm

³ IREON Gateway is the new brand of the International Relations and Area Studies Gateway that was accessible through <http://www.iras-portal.eu>. This IP will be re-directed to <http://www.ireon-gateway.eu>.