Regional studies, regional portals and newspaper content in Germany

Ulrich Hagenah
Hamburg State and University Library Carl von Ossietzky
Hamburg, Germany
Email: u.hagenah@t-online.de

Abstract:

There is a huge variety of digital or digitised source material, inventories, and catalogues in diverse formats on the web relevant for regional studies – but very often hidden in the 'deep web', offered on websites of the producing institutions such as libraries, museums, archives, local and regional public authorities, parliamentary or media documentation services, foto agencies, statistical bureaus etc. Most of the resources are hard to find, and all offer different user interfaces, whereby research is difficult, and the outcome all to often incomplete and insufficient.

German regional portals started to build up focal points for researchers in the past years, bringing together mainly historical sources, and sometimes initiating digitisation efforts with cooperation partners. Most of them are conducted by regional libraries, others by cultural heritage institutions or state archives. Regional portal providers closely cooperate with the institutions producing relevant databases. Some of them also implement elements of crowd sourcing, for instance for biographical databases, photo archives or descriptive local history.

HamburgWissen Digital is the regional portal for Hamburg, being established in a project of Hamburg State and University Library from 2011-2013. All databases containing Hamburg issues as far as they are technically applicable will be merged into a metasearch environment which allows scholars to receive diverse sorts of source material (text, pictures, archival registers, museum objects, statistics, etc.) through only one query. Newspaper content will be a most important part of this body of acquired consolidated knowledge, be it from digitised newspapers or clipping collections. The presentation will stress the newspaper aspect as well as biographical web resources.

What I am going to talk about is a project, based in Hamburg, called *HamburgWissen Digital (HWD)* [1], destined to enhance access to resources relevant for regional studies of all kinds.

---

[1] Literally translated: 'Digital knowledge and data resources about Hamburg' – a proper title in English still has to be found.

The starting point of this project was our observation:
There is a huge variety of highly relevant digital or digitised source material, inventories, and catalogues in diverse formats on the web – but very often it is hidden in the 'deep web', offered only on websites of the producing institutions such as libraries, museums, archives, local and regional public authorities, parliamentary or media documentation services, photo agencies, statistical bureaus, etc. Most of the resources are hard to find, and all of them offer different user interfaces, making research difficult, and the outcome all too often incomplete and insufficient. We would like to change this for our customers.

We: that is the Hamburg State and University Library[2] in cooperation with 20 to 25 cultural heritage institutions from Hamburg and elsewhere. Our library serves as university library for about 40.000 students and as regional library for the federal state of Hamburg, one of 16 German states (Länder). This means: as a state library we have

- legal deposit,
- a comprehensive collection of regional literature, maps and drawings of the city, portraits, bequests of well-known people of the area, manuscripts, etc.
- a regional bibliography
- digitisation projects bringing unique content to the web
- the task to present our Northern German cultural heritage in different ways by means of exhibitions, publications, colloquia, and again of course: websites.

The library now has about 4.3 million holdings, after losing roughly 80 % of its collections in World War II, for example the whole newspaper collection – replaced partly in original, mostly by microfilms from holdings of other institutions. Thus, it seems advisable to cooperate. The consequence in all fields of cultural heritage promotion is: cooperative action, bringing together complementary resources, trying to make the most of the situation which is both difficult but also quite promising.

Hamburg, as the center of a metropolitan area with 1.7 million inhabitants, has various institutions collecting, preserving its cultural heritage and carrying out or initiating research about it: four museums on regional history, a state archive and several universities. The biggest of them, Hamburg University, comprises institutes and collections on regional history, geography, social and economic structures. The State and University Library provides its digitised books on regional history, journals, maps, drawings, portrait collection and other items, for example 2.1 million pages of 7 newspapers being currently digitised and converted into searchable text within the first half of next year. Many other libraries, local and regional public authorities, parliamentary or media documentation services, photo agencies, statistical bureaus etc., research institutes and other corporations and authorities are concerned with the preservation and promotion of our cultural heritage. There are interesting private archives, photo collections and agencies and, most important as the library's customers, groups of local historians with special interest in the urban quarters and districts. Many of them provide important digital collections. Rich documentary resources can be found at the historic buildings and monuments protection authority and in the former HWWA (Hamburg Archive of International Economics) with its vast collection of newspaper clippings on economics, politics and history. And moreover, in other parts of Germany important digital collections of pictures, museum objects or archival material comprise lots of subjects relevant for research in our region.

So we decided to build up a joint portal in order to bring these digital resources together:
1. into a link catalogue, and
2. merge all these resources, as far as they are technically applicable, into a metasearch environment which allows scholars to receive diverse sorts of source material (text, pictures, archival registers, museum objects, statistics, bibliographical and biographical information, etc.) through only one query.
3. We provide additional services such as
- guides to archives and libraries of the region,
- a blog, with the goal to create a cooperative communication and research platform for all people interested in our region,
- a table of events serving as a calendar to all sorts of lectures, guided tours, exhibitions, etc. related to Hamburg topics.
- A tutorial on conducting research in regional history and geography is being prepared for next year.

Our target groups are all academic or non-academic researchers in regional studies:
- scholars and students of the universities
- local and regional historians, geographers, genealogists[3]
- the media
- teachers
- the broad public

Interests, queries and the manner of use may differ between the respective user groups. Scholarly use will be mainly quite specific, aiming to maximise and optimise recall for elaborated research questions, while helping pupils or journalists often means giving a first impression of the variety of sources for very broad themes. The project team is currently working on several use cases representing typical queries and situations of customers. User group specific tests are going to follow.

We were not the first in Germany to have the idea of a regional portal. There are already several of them, for example in Bavaria – the first (since 2002) and leading portal in Germany called "Bayerische Landesbibliothek online"[4], Hesse, Thuringia, Saxony, in the West and Southwest of the country, created as focal points for researchers in the last years, bringing together mainly historical sources, and sometimes initiating digitisation efforts with cooperation partners. While Bavaria or Westphalia[5] concentrate on assembling widespread collections of historical source material – manuscripts, portraits, photos, biographical information, historical journals and regional encyclopedias -, the regional information system of Hesse[6] has its focus on historical maps and drawings, topographical information and atlases. The Southwest German portal "Regionalgeschichte.net"[7] successfully stimulates local historians and groups to publish their resources, sometimes complete local histories of towns and villages on the web, and the portal-makers provide diligent support to local activities in Rhineland-Palatinate and the Saarland.

HamburgWissen Digital as the regional portal for Hamburg and its metropolitan region does not assemble digital content in a kind of data warehouse itself, but bundles all the relevant resources in regional history

---

[3] About 12.500 of the 52.000 customers of the State and University Library are not affiliated with Hamburg University and use it for private purposes.

[4] http://www.bayerische-landesbibliothek-online.de <28.5.2012>. – A joint website of all German regional portals shows their thematic and organisational variety: http://www.lwl.org/westfaelische-geschichte/ag/index.html <28.5.2012>.

[5] http://www.westfaelische-geschichte.lwl.org <28.5.2012>.

[6] http://www.lagis-hessen.de <28.5.2012>.

[7] http://www.regionalgeschichte.net <28.5.2012>.

and regional studies in order to enhance visibility and to facilitate cross-search. The Hamburg based project is the first one (similar to those in Stuttgart in the Southwest for Baden-Württemberg and in Jena for Thuringia) to enter into the field of metasearch engines for regional resource material. Our special purpose is to offer metasearch on a combination of resource types: digital objects, full text, content of factual databases, and digital references or inventories (of libraries, archives, museums and other collections). Combining digital objects with reference services makes the difference of our portal from the object-centered portals like Deutsche Digitale Bibliothek and Europeana.

Most of the German regional portals are conducted by regional libraries, others by cultural heritage institutions or state archives. Regional portal providers closely cooperate with the institutions producing relevant databases. Some of them also implement elements of crowd sourcing, for instance for biographical databases, photo archives or descriptive local history. Newspapers are part of the Bavarian portal and will play a more visible role in the portals as soon as newspaper digitisation will have proceeded further and gained more results than up to now. Newspaper content will be a most important part of this body of acquired consolidated knowledge, be it from digitised newspapers or clipping collections. In our case there is a large clipping collection of a former institute on economic research, and we will integrate the collection of digitised newspapers of our library as soon as it will be published at the beginning of 2013 and full-text-searchable.

HamburgWissen Digital was established as a project of Hamburg State and University Library at the beginning of 2011 and will last for probably two and a half years in order to be continued as part of the library's services in cooperation with external partners. What we can offer at the moment looks like this:

- homepage www.hamburgwissen-digital.de with the simple search field, the display of thematical clusters of the link catalog and the button "Weitere Angebote" which leads to several services as our guide to archives, our blog or the event calendar.

- <u>the link catalog</u>, here the example of local history: it provides comments on the range of coverage, search tips, advice on related resources, information about the points of media break and the change to undigitised paper material
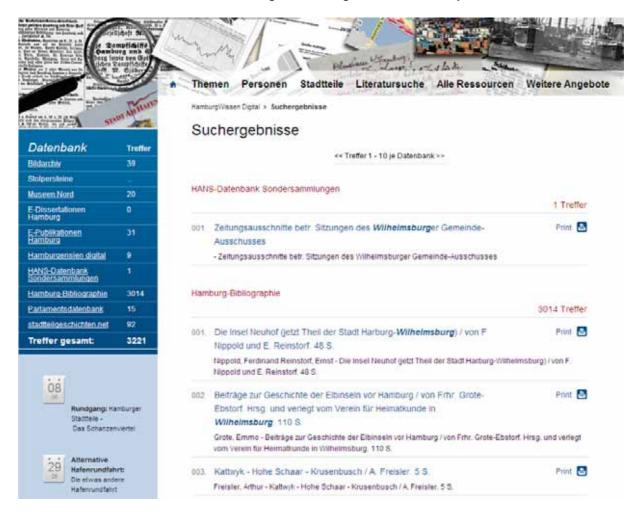

- the <u>metasearch</u> with simple and extended search interfaces.

Let me mention just some crucial points of our work on the metasearch environment, the centrepiece of *HamburgWissen Digital*:
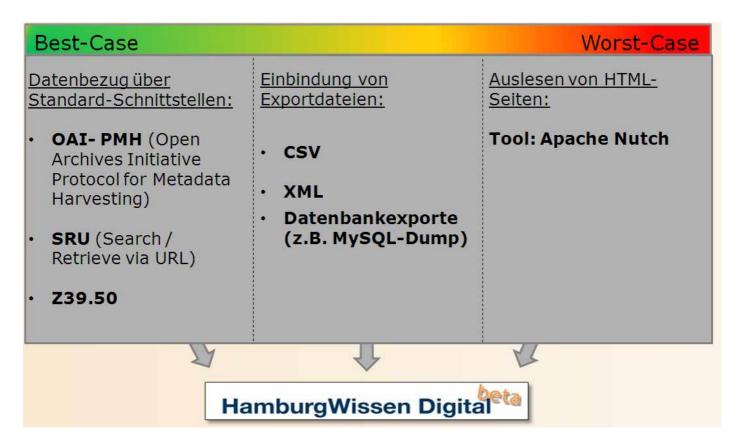
And here are search results in *HamburgWissen Digital beta,* listed by source databases.

Building up the metasearch environment firstly requires a detailed analysis of the – probably about 50 – source data records to be embedded, followed by the conceptual and technical mapping of all their content categories on 10 defined metasearch fields. For every source database software gateways have to be created within our portal software IPS (Information Portal Suite) in order to address the search correctly and receive adequate document delivery.
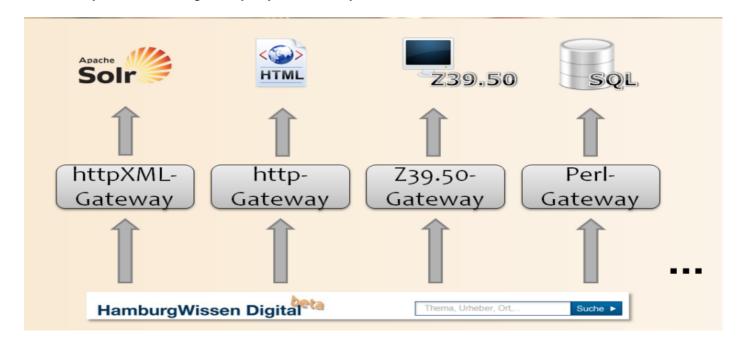
The metasearch engine in principle does not ingest and manipulate the source data themselves, but appeals to the search functionalities of the different partner institutions and presents their results as a standardised hit list. For every connected partner resource a so-called gateway is furnished in the HWD search engine which translates the inquiry of the HWD metasearch into an inquiry of the respective partner searching machine.

The source data of our partner institutions on the web are accessible in different ways. There is a broad range, at best from precisely defined interfaces up to the worst case of data that are not available in a well-defined format, so that metadata have to be parsed from the HTML pages of the supplier.

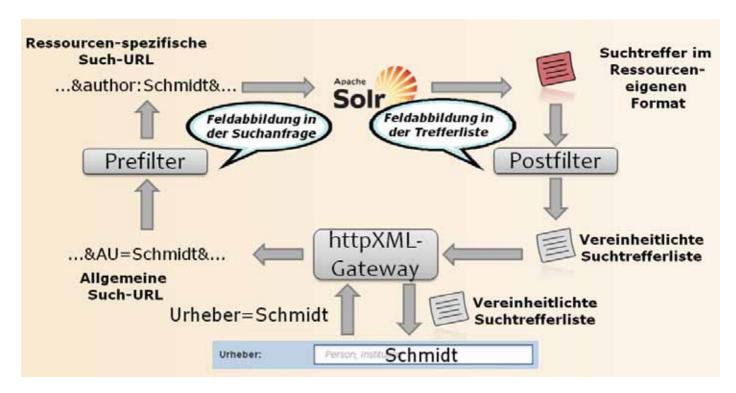| Best-Case		Worst-Case
Datenbezug über Standard-Schnittstellen:	Einbindung von Exportdateien:	Auslesen von HTML-Seiten:
• **OAI- PMH** (Open Archives Initiative Protocol for Metadata Harvesting) • **SRU** (Search / Retrieve via URL) • **Z39.50**	• **CSV** • **XML** • **Datenbankexporte (z.B. MySQL-Dump)**	**Tool: Apache Nutch**

HamburgWissen Digital ^beta

In the optimum case the partners can offer a standardised interface, as for example OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting). Then HWD uses this interface to ingest the metadata and transfer them into HWD's own Solr index. There are also data sources of cooperation partners that do not offer any search interface on the web, so that no inquiry can be defined via URL parameters. In this case data selections (in CSV, XML formats, or as a MySQL dump) are transferred to the HWD Solr index in order to integrate these sources nevertheless. In case we cannot receive results in a structured format like XML, the relevant content has to be selected from HTML pages with the help of technologies like Apache Nutch. However, we try to avoid this version because owing to changes in the HTML pages the selections have to be adapted over and over again.

For the very different data resources we need a highly flexible IT system. A combination of the systems IPS and Apache Solr serves as our technological basis of the metasearch. IPS delivers the metasearch functionality in the form of gateways by which every data resource is connected.
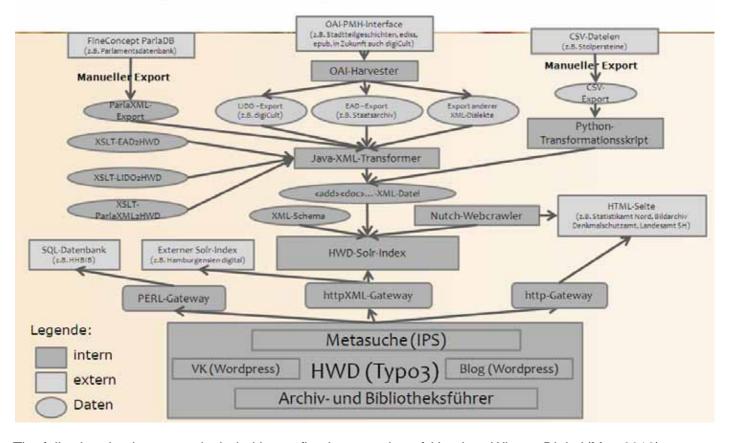
The gateways operate the different forms of the query, and the search technology Apache Solr offers the possibility to make data accessible very quickly and, in a later stage of the project, to sort the results according to defined criteria as for example relevance, etc. Here you see the data flow for a query about an author named Schmidt:

By now the project has reached the technical status that all conceivable format variations of the partner resources can technologically be handled, so that new data resources can be integrated on the same basis in the future. The following diagram gives a synopsis of all technological approaches in the project structure.

HamburgWissen Digital

The following databases are included in our first beta version of *HamburgWissen Digital* (May 2012):

HWD-Metasuche – Ressourcen aktuell

Status	Datenquelle	Institution
🟢	**Bildarchiv des Denkmalschutzamts**	*Denkmalschutzamt Hamburg*
🟢	**Parlamentsdatenbank 1997-**	*Hamburgische Bürgerschaft*
🟢	**stadtteilgeschichten.net**	*stadtteilgeschichten.net e.V.*
🟢	**Stolpersteine** [Biographien]	*Landeszentrale für politische Bildung (HH)*
🟢	**Museen Nord** [zunächst: Hamburger Museen]	*digiCULT-Verbund eG*
🟢	**Hamburgensien digital**	*Staats- und Universitätsbibliothek Hamburg*
🟢	**HANS-Datenbank Sondersammlungen**	*Staats- und Universitätsbibliothek Hamburg*
🟢	**Hamburg-Bibliographie / Hamburger Personen**	*Staats- und Universitätsbibliothek Hamburg*
🟢	**E-Dissertationen Hamburg**	*Staats- und Universitätsbibliothek Hamburg*
🟢	**E-Publikationen Hamburg**	*Staats- und Universitätsbibliothek Hamburg*

And the following resources are to be integrated within June/July 2012:

HWD-Metasuche – Ressourcen in Vorbereitung

Status	Datenquelle	Institution
🟡	**Zeitungsausschnittarchiv des ehem. HWWA**	*ZBW Hamburg*
🟡	**Parlamentsspiegel 1980-**	*Hamburgische Bürgerschaft / LT NRW*
🟡	**Statistik**	*Statistikamt Nord*
🟡	**Das Jüdische Hamburg** [Biographien, Themen]	*Institut für die Geschichte der dt. Juden*
🟡	**Landesrecht Hamburg**	*Justizbehörde Hamburg / Juris*
🟡	**Rechtsprechung Hamburg** („Urteilsdatenbank")	*Justizbehörde Hamburg / Juris*
🟡	**Stiftungsdatenbank**	*Justizbehörde Hamburg*
🟡	**AIDA online – Findmittel LA Niedersachsen**	*Landesarchiv Niedersachsen*
🟡	**Handels-, Firmen-, Genossenschaftsregister**	*Justizbehörde Hamburg*
	.....	

There are several among these resources comprising much more than only Hamburg-related content – so we have to extract the relevant material, using a kind of "Hamburg filter": simply spoken, we select relevant documents and objects which contain Hamburg-specific vocabulary in their metadata: geographic terms, personal or corporate names, specific terms related to our city.

So, to sum up, I would like to stress the services that our project is going to offer:
- bringing resources forward to users
- enhancing visibility of certain resources
- facilitating the research process, the comparison and combination of resources
- giving orientation, a communication platform and discussion forum for regional studies topics

http://www.hamburgwissen-digital.de/home.html: this is our website where you can observe the progress of our project, if you like – we would be grateful for your comments and advice. Thank you.