

Sondera - a portal for serendipitous discovery of cultural heritage

Anders Söderbäck, National Library of Sweden

Abstract

Sondera (<http://sondera.kb.se>) is a Swedish search portal developed in cooperation between the National Library and the National Archive. The portal makes it possible to discover various types of material - books, journals, TV and radio programs, archival collections - from three national sources: The national library catalog, the national archival database, and the Swedish media database.

The purpose of Sondera is to provide a broad entry point to Swedish cultural heritage. The focus during development was in quick serendipitous discovery rather than in-depth research. This presentation will give an introduction to why and how the portal was designed and to the principles that guided its development.

Background

In 2007, discussions were being held between three Swedish government agencies; the National Library, the National Archive and the National Archive of Sound and Moving Images, about joint projects for working together and learning from each other, with the long term goal of better interoperability between technical systems. The main project that was born out of these discussions was to be named "samsökningsprojektet" (best translated into English as "the portal project"). The purpose of this project was to provide a common entry point for searching the three big databases that the three agencies were responsible for: LIBRIS, NAD and SMDB.

The three databases are openly accessible and are used by researchers and scientists as well as by the general public. For the three agencies, it was from the beginning obvious that the databases could not easily be replaced by a single service or search interface. In the discussions that preceded the project it also became clear that the different data models being used by the three databases differed so much that, if something was to come out of a joint project within reasonable time, the three databases could not be integrated on any deeper level.

Instead of putting a lot of work into discussing bibliographic format and the different traditions of data modeling in libraries and archives, it was decided that the project should be technically very light weight. The common portal should use the SRU protocol for searching and retrieving information from the three databases. This information was to be presented in a common interface. For exhaustive presentations of the individual resources, the service would provide deep links back to the native databases. Instead of infrastructure and data integration, this project would focus on good design and on providing a different experience than what could be provided by the native databases individually.

The project begun autumn 2008 all three institutions working together. On January 1st 2009, the National Library and the National Archive for Sound and Moving Images was merged into one organization. This did not affect the project and Sondera was launched on April 1st 2009.

Infrastructure

To cover the technical requirements it was necessary to provide SRU-interfaces to all three databases. LIBRIS already had an openly available SRU interface, and setting one up for SMDB was easily accomplished. Providing SRU access to NAD proved to be a little more difficult. Due to performance issues in the search engine being used by for NAD, search and retrieval was not fast enough to be used in Sondera. The solution was to store a copy of NAD at the National Library and make use of the search engine powering LIBRIS to provide SRU access to NAD. This copy of NAD is being kept up to date using the Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH.

To avoid putting too much work into metadata conversion, the service was to rely mainly by keyword searching. However, three possible refined access points was implemented: title, creator and year.

Design

A lot of work was put into interviewing users and stakeholders as well as into prototyping and testing the service. A key strategy for the project was working together. One aspect of this

strategy was to get people from all three organizations involved in all parts of the project. Another aspect was to have all parts of the team working together on all parts of the service to provide good communication and sharing of ideas between developers, designers and librarians. This proved a good strategy. The project was finished on schedule and was well received.

Instead on giving an exhaustive presentation of the collections, the main objective for Sondera was to provide a quick and easy outline or overview of different resources represented in the three native databases. Providing such an overview seemed logical, given that the three databases would contain for example books by, letters written to, and tv-documentaries on a certain author. This approach also meant that the material categories being put in the catalog records but not always that visible in traditional search services was given high priority in the Sondera interface.

Getting such an overview of the collections was received very well in user testing of the service. In fact, the most appreciated feature of Sondera seemed to be the possibilities of serendipious discovery, of finding resources you didn't even know that you were looking for. The interface for Sondera was designed to be intuitive and easily understandable, but also to encourage curiosity and excitement. This key aspect of the service also made it's way to the actual name of the service. "Sondera" in Swedish might be translated as "probing" or "scanning".

Access to content

The biggest problem with building Sondera was the question of providing access to actual content, or rather how to easily explain to end users how to get such access. All content being represented in LIBRIS, NAD and SMDB is publically available, though usually not easily available and definitely not available online. The physical holdings in LIBRIS is accessible through the libraries directly or as interlibrary loans, archival collections usually demands a visit to the archive and to access the audiovisual content in SMDB it is necessary to visit the National Library or one of the currently 41 institutions providing access to audiovisual collections through interlibrary loan.

Digitized or electronic content is accessible online, though not always openly accessible - access to libraries licensed content need to be mediated by the library. A quick and easy overview of

content without a quick and easy way to access the content often frustrated end users, and a lot of effort was put both into providing good explanations of where to access the actual content and into highlighting content that was actually openly accessible.

Discussion

No formal evaluation of the use of Sondera has been done post project. However, it is clear that the usage of Sondera is significantly smaller than the usage of the three native databases. During 2009, Sondera was had an average of 6 200 visitors per month performing an average of 15 000 searches per month. This is only a tiny fragment of the 14 million searches that was handled by LIBRIS during the same year. The low usage of Sondera could be attributed to lack of marketing - a user not knowing about Sondera will not use the service. Sondera is now being marketed using Google AdWords. It is, however, too early to determine if this has affected usage of the service.

Another possible explanation for the low usage of Sondera is that the potential user base for the service might actually be significantly smaller than the user base for the native databases. Combining information from LIBRIS, NAD, and SMDB does not necessarily imply a user base containing anyone interested in library, archive or audiovisual resources. It might be argued that the user base for such a service would be persons interested in all three native databases. A student or scholar in the humanities might be interested in library collections, a historian in the archival collections and a media historian in the audiovisual resources. The ideal Sondera user might be a student or scholar combining an interest in the humanities with an interest in both history and media history. Preliminary evaluation of the use of Sondera indicates that usage comes from the archive sector rather than the library sector. This does not imply any failure on part of Sondera. Rather, it means that Sondera is a niche service being aimed at a smaller audience. Such a niche service will probably never achieve mainstream status, but might still be highly appreciated by the people using it.

The possibility for building niche services might be discussed as being part of the trend of decoupling the front end of the systems from the back end. This decoupling has been described by Marshall Breeding as being "[o]ne of the seminal breakthroughs in library automation

involves the separation of resource management from resource discovery.” In Breedings interpretation, this trend suggests the possibility for unifying resources originating in different administrative systems into fewer interfaces, ”hopefully culminating in a single entry point to all library content.” The decoupling of infrastructure and interface, however, also gives the opportunity of using resources from the same administrative system in several interfaces, each targeted for specific audiences. When viewed from this perspective, Sondera is successful both in providing a good experience for a niche audience and was also successful in giving the institutions participating in the project valuable experience in creating a service that fulfills different user needs than what the traditional library or archive services do.

However, this does not answer the fundamental problem of how to provide access to the actual content. Despite the efforts being put into Sondera, this remains an issue that could not be solved by this project. Rather, this is something that has to be solved by putting more efforts into digitization, open access publishing and licensing.