

Lund University Libraries

National Reporting System for Library Statistics in Sweden

Investigation of existing national solutions in Denmark, Finland, Norway and Sweden plus description of proposed new national system in Sweden

Poul Henrik Jørgensen, PHJ @ Portia.dk
22-12-2010

Introduction

This report was prepared in response to an initiative by the Swedish Royal Library Expert Group for Library Statistics in 2010.

The work was managed by Lund University Libraries represented by Deputy University Librarian Tore Torngren with funding contributed by the Swedish Royal Library.

The work was carried out by IT Consultant Poul Henrik Jørgensen who prepared the first version of this report in October 2010.

The present version 2f was completed at the end of December 2010.

Contents

Introduction.....	2
Summary.....	6
Abstract	9
1 Background and Objective.....	10
2 Preliminary Analysis.....	12
2.1 User Requirements.....	12
2.2 System Functions.....	13
3 Analysis of existing national Solutions	16
4 Library Statistics in Denmark.....	17
4.1 Organization in Denmark.....	17
4.2 Information Collected in Denmark	17
4.3 Distribution of Information in Denmark.....	18
4.4 Technical Solutions in Denmark	19
4.5 Systems Development and Operation in Denmark.....	19
4.6 Current Plans in Denmark.....	20
4.7 Conclusions from Denmark	20
5 Library Statistics in Finland	21
5.1 Organization in Finland.....	21
5.2 Information Collected in Finland	21
5.3 Distribution of Information in Finland.....	22
5.4 Technical Solutions in Finland	22
5.5 Systems Development and Operation in Finland.....	23
5.6 Current Plans in Finland.....	24
5.7 Conclusions from Finland	24
6 Library Statistics in Norway	26
6.1 Organization in Norway.....	26
6.2 Information Collected in Norway	27
6.3 Distribution of Information in Norway.....	27
6.4 Technical Solutions in Norway.....	27
6.5 Systems Development and Operation in Norway	28
6.6 Current Plans in Norway.....	28

National Reporting System for Library Statistics in Sweden

6.7	Conclusions from Norway	28
7	Library Statistics in Sweden	29
7.1	Organization in Sweden.....	29
7.2	Information Collected in Sweden	29
7.3	Distribution of Information in Sweden.....	30
7.4	Technical Solutions in Sweden	30
7.5	Systems Development and Operation in Sweden	30
7.6	Current Plans in Sweden.....	31
7.7	Conclusions from Sweden	31
8	Lessons from existing National Solutions.....	32
8.1	Organization	32
8.2	Information Collected.....	33
8.3	Distribution of Information	33
8.4	Technical Solutions	34
8.5	Systems Development and Operation.....	34
8.6	Current Plans	34
9	General System Requirements	35
9.1	Evolving Data Models	35
9.2	Diverse Data Sources	35
9.3	Different Dissemination Methods	35
9.4	Customization Facilities	36
9.5	Self-service Analysis Facilities.....	36
9.6	Data Management.....	36
9.7	System Deployment	37
10	Implementation Strategy.....	38
10.1	Build on Standard System.....	38
10.2	Deploy Central Shared Services.....	38
10.3	Develop Iterative and Incremental Solutions.....	39
11	Standard Products and Services	40
12	Proposed Facilities and Services.....	41
12.1	Survey Management	41
12.2	Data Acquisition.....	41
12.3	Data Management.....	41

National Reporting System for Library Statistics in Sweden

12.4	Information Dissemination.....	42
13	Prototype Examples.....	43
14	Notes	46

Summary

This report describes the proposed technical solution for a new national library reporting system for library statistics in Sweden. The proposed solution is based on a preliminary requirements analysis supplemented by an investigation of the existing national library reporting systems in Sweden plus Denmark, Finland and Norway.

Functional requirements to a new national system for library reporting in Sweden were analyzed during a preceding project. General non-functional system requirements were not addressed by that project, because the future organization and management of the system were unknown at the time – and still is at the time of this writing.

The current project undertook to investigate existing systems for national library statistics within Denmark, Finland, Norway and Sweden. Results and lessons learned from these existing solutions were analyzed to identify relevant general requirements to the future system in Sweden.

The other investigated countries use proprietary integrated systems to handle their respective library statistics on a national level. Sweden is alone in adopting a more loosely structured solution built on standard commercial products. The Swedish system has been able to handle evolving requirements by means of more informal ongoing adaptations.

The user requirements analysis together with lessons learned from the existing solutions demonstrate that the national reporting system for libraries must be designed to handle evolving data models and data sources- including transactional data extracted from other systems.

In order to satisfy this requirement for flexibility and versatility, it is proposed to take a lead from the existing Swedish system and base the new solution on a general commercial system for data management and analysis – instead of implementing a highly structured proprietary application from the bottom up.

The proposed strategy will allow Sweden to establish a prototype system that can quickly be put into service. The prototype system may gradually evolve into a fully operational solution that can satisfy relevant current and future requirements.

This approach is more efficient and less risky than a more traditional implementation strategy, which is better suited to situations where the data schema and business rules are relatively static and well-defined.

The following diagram illustrates how the proposed solution can support different scenarios:

- Some libraries may choose to submit their statistical data directly to the central system
- Other library organizations may prefer to pre-process their statistical data within distributed systems established through regional or organizational cooperation.
- Data from libraries is imported to the central systems where it may be combined with information imported from other external systems
- Information can be published directly from the central system in the shape of reports or diagrams
- Selected information can also be exported to other external systems and databases

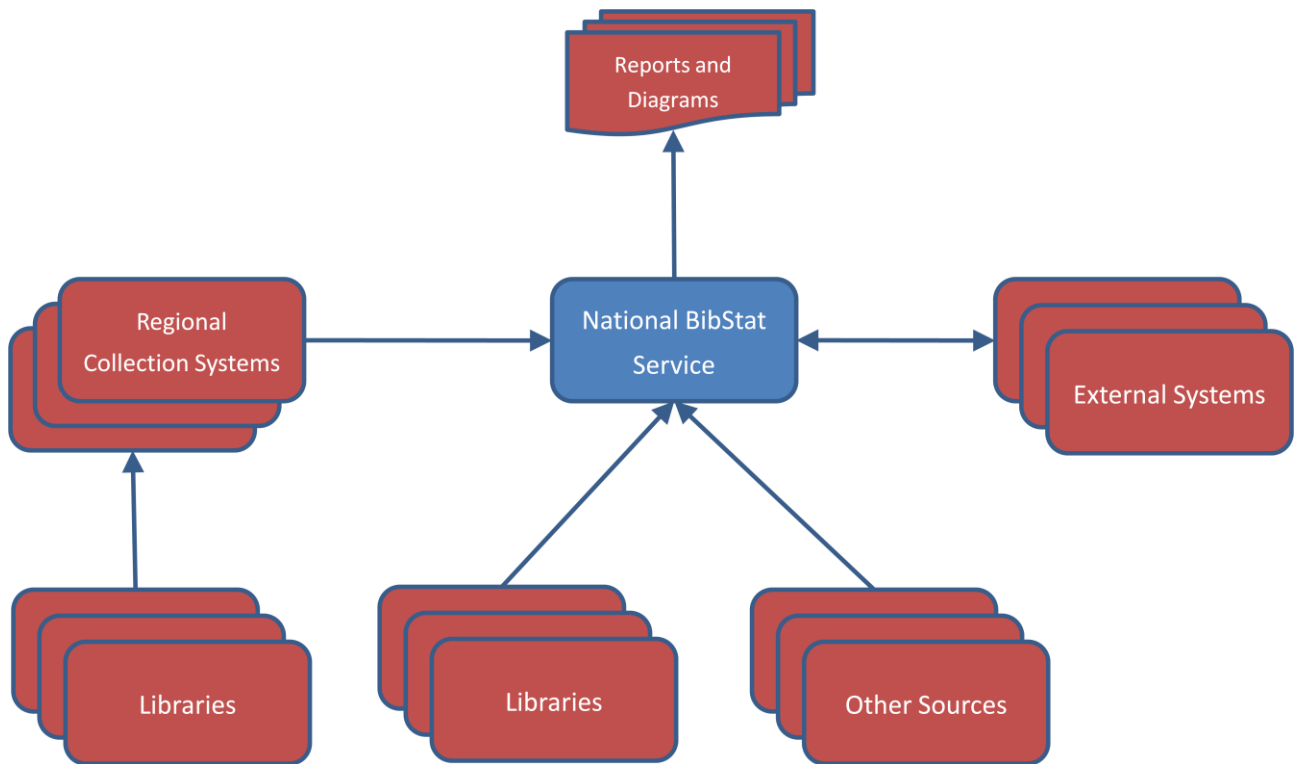


Fig. 1 Scenarios for external dataflow

The following diagram illustrates how the proposed solution can import statistical data in different formats including files with spreadsheets or XML documents.

The system can also import data directly from external systems using a variety of widespread database systems. The central system can maintain dynamic links to some types of external databases etc., so that changes to linked records within the local databases are automatically reflected within the central system.

The proposed system can export statistical information as printable reports, spreadsheets, XML documents and special files to external systems.

The solution also includes standard service interfaces, so that designated external systems can fetch data on-line via system-to-systems web services.

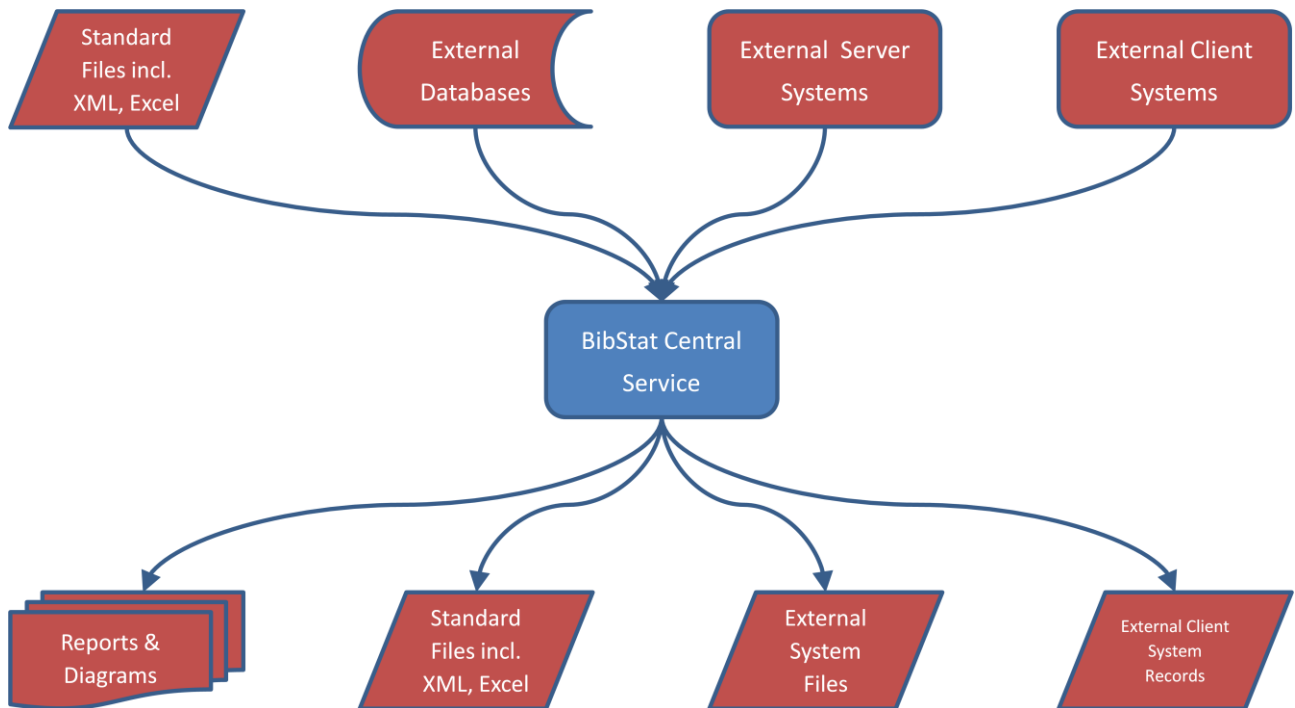


Fig. 2 Types of external data exchange

The national system includes general self-service tools, so that libraries and cooperating organizations can implement their own customized self-service solutions to collect and analyze statistical data.

Some of the most demanding features of the proposed solution have been implemented by an informal prototype in order to demonstrate, that the system can in fact satisfy the demanding functional requirements. The prototype tests also indicate that the solution is scalable to handle the expected performance requirements.

Abstract

The primary objective of this report is to identify requirements to the new national reporting system for libraries in Sweden and to describe a technical solution, which can satisfy those requirements.

The report includes the following chapters:

- **Summary**
- **Background.** Describes the general background, organization and objectives of this work.
- **Objectives.** Describes the specific objectives and methodology of the present work.
- **Preliminary Analysis.** Summarizes the functional user requirements and describes the logical system architecture designed to satisfy those requirements.
- **Existing National Solutions.** Describes the methodology used to investigate existing national system for library statistics in Denmark, Finland, Norway and Denmark. The following aspects are described in relation to the current national solutions within each of the four countries:
 - Organization
 - Data Acquisition
 - Information Dissemination
 - Technical Solutions
 - Systems Development
 - Current Development Plans
 - Conclusions and Recommendations.
- **Lessons from Existing National Systems.** Summary of most important results from the analysis of existing solutions in the selected four Nordic countries. These results have a direct impact on the general system requirements and the proposed solution for Sweden. The lessons learned from the existing solutions are summarized within the following areas:
 - Organization
 - Information Collected
 - Distribution of Information
 - Technical Solutions
 - Systems Development and Operation
 - Current Plans
- **General System Requirements.** Describes general system requirements derived from the investigation of existing national solutions.
- **Implementation Strategy.** Describes the proposed solution and implementation strategy corresponding to functional requirements and system architecture identified by the present work.
- **Standard Products and Services.** Specifies commercial products that are proposed as the basis of the new national reporting system for libraries in Sweden.
- **Facilities and Services.** Describes facilities and services provided by the new system in relation to the specified user functions.
- **Prototype Examples.** Presents a limited functional prototype designed to demonstrate some of the advanced features of the proposed solution. The prototype integrates existing data from the national Swedish library statistics together with other relevant sources.

1 Background and Objective

The *National Library of Sweden Expert Group on Library Statistics*¹ has initiated a number of activities to facilitate a new National Reporting System for Library Statistics in Sweden.

Based on a 2008 proposal from the Expert Group, the *National Library of Sweden*² together with the *Swedish Arts Council*³ commissioned a project in 2009 in order to identify functional user requirements to the new national reporting system. This project was managed by Tore Torngren from *Lund University Libraries*⁴ and the work was carried out by IT Consultant Poul Henrik Jørgensen with funding provided by the Arts Council and the Royal Library.

User requirements were collected from selected library representatives via personal interviews and surveys with interactive web questionnaires. Participants were asked to prioritize a wide range of general user functions and could also suggest any additional functions to be considered.

The user requirements survey revealed a fairly high degree of consensus among the participating library representatives. It was therefore possible to specify the relevant functional user requirements in relation to abstract system architecture with reasonably few designated system functions and interfaces.

During the same time as the user requirements project, the Swedish government announced plans to review the organization and management of national library statistics in Sweden. It was therefore difficult to address technical systems requirements to the national reporting system for libraries in Sweden, because the future organization and management of the system was unknown at the time of writing.

The results of the user requirements project are documented in the report titled *Functional Requirements to Swedish Library Reporting System*⁵ published by the Swedish Arts Council in November 2009.

In order to proceed despite the uncertainties about the future organization and management of national library statistics, the National Library of Sweden Expert Group for Library Statistics approved a two-stage proposal:

- The first task would be to investigate existing systems in Denmark, Finland, Norway and Sweden in order to identify relevant solutions and experiences.
- The lessons learned from these existing solutions together with the functional user requirements would be used to identify systems requirements for the future national reporting system in Sweden.

The National Library of Sweden agreed to provide funding for this project to be managed by Lund University Library represented by Tore Torngren.

Lund University Libraries Head Office commissioned Poul Henrik Jørgensen to execute the project as a natural continuation of the preceding User Requirements Analysis.

This report is therefore a logical follow-on to the preliminary analysis of the future national reporting system for libraries in Sweden.

National Reporting System for Library Statistics in Sweden

The primary objective of the present report is to describe a general system design that can be used to plan prototype tests and subsequent implementation of the new national system for library reporting in Sweden.

While the functional user requirements are primarily based on the results of the preliminary analysis project; the technical system requirements are based on lessons learned from the investigation of existing solutions in the Nordic countries.

Detailed design specifications are beyond the scope of the present project. The current objective is to establish a technical basis for planning of prototype tests and development of the new national system for acquisition, management and dissemination of library statistics in Sweden.

2 Preliminary Analysis

2.1 User Requirements

During the foregoing preliminary analysis, user requirements to the new national reporting system for libraries in Sweden were collected through a series of meetings and web surveys involving officials representing different types of libraries; i.e. public municipal libraries, academic and special libraries, hospital libraries and school libraries.

A list of nineteen user functions was drawn up based on preliminary analysis and inputs from a few potential users. The participating library representatives evaluated the proposed functions and assigned a relative priority to each function.

The list of user functions together with counts of the users' designated priorities of *High*, *Medium* or *Low* is tabulated below.

E.g. 18 representatives assigned a *High* priority to the first function listed in the following table. Seven representatives assigned *Medium* priority and three gave it a *Low* priority. None of the participants expressed *No Opinion* (indicated by question mark) about this function.

No	Functions	H	M	L	?
f1	Proprietary types of information: Individual libraries can submit proprietary types of statistical information (e.g. number and types of patrons attending different public activities).	18	7	3	0
f2	Manual data entry: Library staff and other data contributors can enter a limited amount of statistical information via a central web application.	14	9	3	2
f3	Shared types of transactions: National authorities can collect large datasets with shared types of transactions extracted from operational library systems (e.g. Interlibrary Loan requests).	20	6	0	2
f4	Proprietary types of transactions: Libraries can submit large data sets with proprietary types of transactions extracted from operational systems (e.g. anonymized circulation transactions).	15	12	0	1
f5	Manual upload of data: Data contributors can manually upload datasets to a central web application.	14	9	4	1
f6	Upload via web service: Customized local applications can automatically submit data to a central web service.	20	7	0	1
f7	Xml documents and schemas: Submitted datasets must contain xml documents that comply with designated XML schemas (XSD).	11	3	1	13
f8	Metadata about datasets must include information about the origin, ownership, creation date and content plus reference to the types of data objects contained within the datasets.	20	3	0	5
f9	Metadata about data objects must include information about the structure, meaning and validation rules for the data objects.	13	7	0	8
f10	Semantic relations: Metadata must include information about semantic relationships between data elements (e.g. hierarchy and synonyms).	14	8	0	6
f11	Metadata maintenance: Data contributors can maintain their metadata via a central web application.	19	6	0	3

National Reporting System for Library Statistics in Sweden

f12	Data maintenance: Data contributors can maintain data objects via a central web application.	23	3	0	2
f13	Data Dictionary search: Library staff and other users may search the data dictionary (metadata) via a central web application.	24	3	0	1
f14	Open Access to data: Access to data objects must generally be unrestricted, but not necessarily gratis.	18	1	5	4
f15	Restricted access to data: Access to certain data objects may be limited to authorized users.	4	12	7	5
f16	Data selection: Users can select data objects by means of a central web application with an intuitive search facility.	23	2	0	3
f17	Data reports: Users can present selected data objects via standard reports produced via a central web application.	25	1	0	2
f18	Manual download of data: Users can download selected data via a central web application.	22	5	0	1
f19	Download via web service: Customized local applications can automatically fetch designated data objects from a central web service, e.g. via the Atom Publishing Protocol.	16	5	3	4

2.2 System Functions

Based on the user requirements and external constraints such as existing data models and established standards, a set of necessary general system functions were identified during the preliminary analysis.

The following overview illustrates the proposed major features and logical information flows within the future national reporting system for libraries in Sweden. This does not necessarily reflect the internal system architecture or distribution of functions.

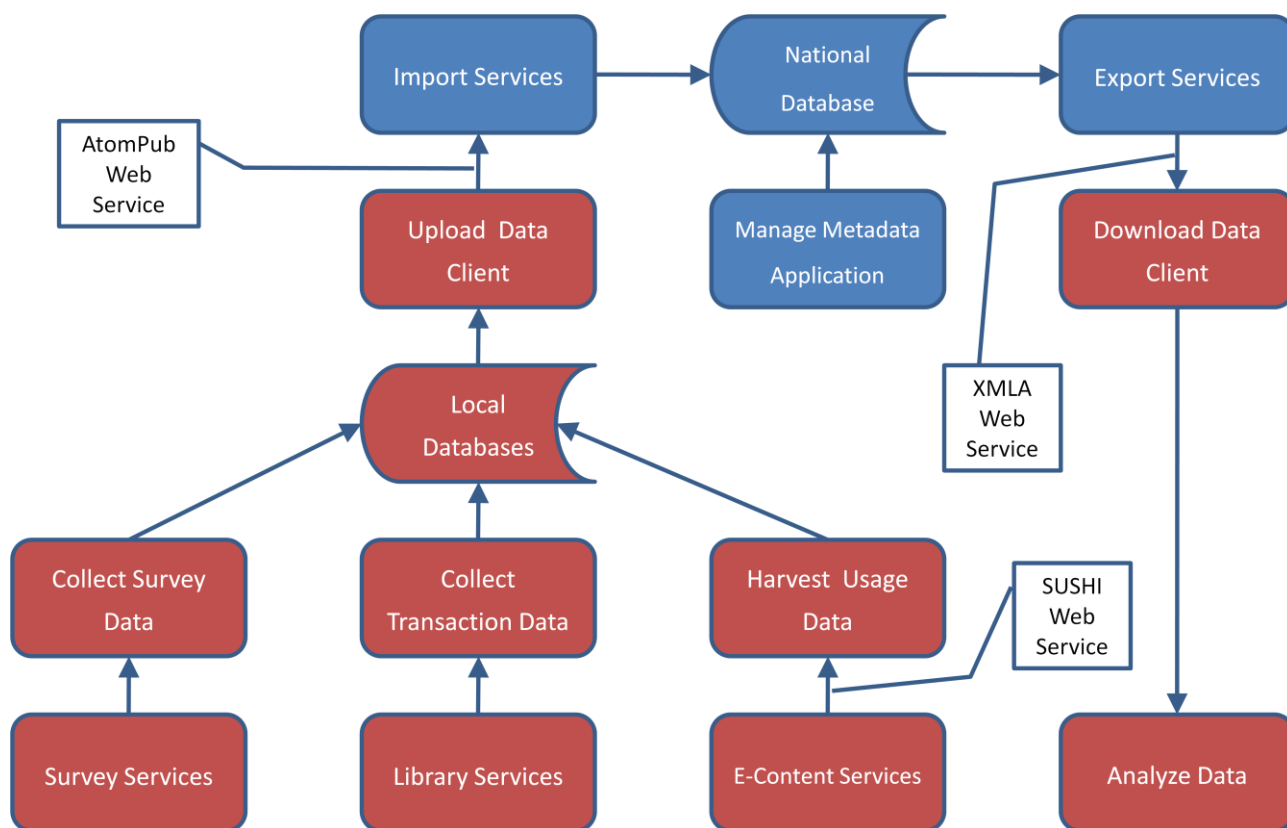


Fig. 3 Logical System Architecture

- **National Database.** The central database is used to manage all of the national library statistics including different types of statistical indicators together with metadata about origins, attributes and logical relationships. The central database also provides facilities to support large scale data mining and analysis.
- **Metadata.** Metadata may describe data structures, domains, attributes and logical relationships as well as information about the authorship, origins and methodology used to collect the data.
- **Collect Survey Data.** Statistical data can be collected via interactive web surveys with functions to validate and manage responses.
- **Collect Local Data.** Statistical information and transaction data can be exported from relevant local library systems.
- **Harvest Usage Data.** Usage data can be harvested from Electronic Content Providers via standard local web services. E.g. COUNTER⁶ type data fetched from SUSHI⁷ compliant services.
- **Local Databases.** Data, that is collected from surveys and local systems or harvested from remote services, can be compiled within local databases. The local databases can act as staging posts, where statistical information and transaction data etc. is verified and consolidated before upload to the national database. Libraries and other contributing partners may use any kind of local database ranging from spreadsheet files to general relational database systems.
- **Data Import.** Data from local systems or survey services etc. can be uploaded to the central Import Service, which store the data within the central database.

National Reporting System for Library Statistics in Sweden

- **Data Export.** Data can be exported from the central database to central analysis services or downloaded to local systems.
- **Data Analysis.** Consolidated information can be presented on-line as standard reports, statistical tables and graphs.
- **Data Download.** Selected data can be downloaded as spreadsheet tables to local analysis and publishing tools.

General non-functional system requirements and implementation methods were not addressed during the preliminary analysis. This was partly because, in September 2009 the Swedish government proposed the establishment of a new independent authority with responsibility for analysis and evaluation of cultural activities from 2011. The future organization and management of national library statistics in Sweden were therefore unknown at the time.

General system requirements are instead derived from analysis of existing solutions for national library statistics within Finland and the Scandinavian counties and Finland.

3 Analysis of existing national Solutions

In order to learn about existing national systems for national library statistics, contacts were established with representatives of relevant organizations in Denmark, Finland, Norway and Sweden.

All of the representatives contacted proved very helpful and set up local meetings in Copenhagen, Helsinki, Oslo and Stockholm where representatives of the organizations responsible for different parts of the respective national solutions met with Poul Henrik Jørgensen.

Before the meetings, the following list of questions was forwarded to the representatives together with some background information about the objectives of the project:

1. How is the collection of national library statistics organized in your country?
2. How is the storage and distribution of national statistical data about libraries organized?
3. What kinds of statistical information about libraries are collected on a national level?
4. Do you collect the same types of statistical data from all of the libraries – or perhaps different types of data from different types of libraries?
5. What is the approximate volume of statistical data about libraries, which has been accumulated on a national level?
6. Do you collect transactional data from operational library systems – if so, how?
7. What kind of technical solutions and IT services are used to collect, manage and distribute statistical library data on a national level in your country?
8. How are statistical data about libraries distributed and published in your country?
9. How are the operation, maintenance and development of the technical solutions and infrastructure organized?
10. Which public and/or proprietary standards are used in connection with the collection and distribution of statistical data from libraries in your country?
11. What kinds of metadata are used to document the statistical data from libraries in your country?
12. Are there any current plans to develop, change or replace the current technical solutions?
13. What advice or experiences would you like to emphasize in relation to the establishment of a new national system in Sweden?

During the meetings these were used as an informal checklist of the issues to be addressed. The participants attempted to address all of the relevant issues.

After each meeting, Poul Henrik Jørgensen prepared a draft summary with relevant information about the respective national solutions. The draft documents were forwarded to the local meeting participants for review. Each of the following descriptions of national solutions has therefore been checked by local representatives, who patiently commented on successive versions of the documents.

Information about the existing solutions has been very helpful to the analysis of general system requirements for the new national reporting system in Sweden. The assistance and information provided by the representatives from Denmark, Finland, Norway and Sweden is greatly appreciated.

4 Library Statistics in Denmark

The *Danish Agency for Libraries and Media*⁸ represented by Leif Andresen hosted a meeting in Copenhagen on 14. September 2010.

4.1 Organization in Denmark

Statistical information about public libraries and research libraries in Denmark is managed by the Danish Agency for Libraries and Media under the Danish Ministry of Culture.

The Ministry of Culture collects annual information directly from the 97 Danish municipalities with public library systems. Information about individual branch libraries is included in the aggregated reporting from each commune.

The Ministry of Culture also collects annual information directly from approximately 50 large research libraries plus approximately 250 smaller ones.

Statistical Information about public school libraries is managed by an agency of the Danish Ministry of Education. Some combined libraries aggregate⁹ the information for school- and public library loans for technical reasons. In such cases, the public library is asked to estimate how much of the total, that should be attributed to the school library. These estimates are included in the school library statistics compiled by the Ministry of Education.

National statistics about public libraries and relevant research libraries are published by the Ministry of Culture¹⁰ while National statistics about public school libraries are published by the Ministry of Education¹¹.

Summaries of the statistical information for libraries are also included in national statistics published by Statistics Denmark¹².

4.2 Information Collected in Denmark

The information collected from Danish libraries is primarily based on the *ISO 2789 International Library Statistics*¹³ standard with some additional levels of details¹⁴.

Different sets of data are collected from public- and research libraries respectively, although both groups share many common data elements. Larger research libraries with more than 20 employees are required to report more detailed information than smaller ones.

The types of data elements vary a little from year to year because some data elements may be dropped and new ones added to the annual collections.

Public libraries submit large amounts of detailed information about loans and holdings in the shape of standardized data files extracted automatically from their operational library applications. Each of the 97 municipalities in Denmark submitted an average of 7.200 statistical data elements for 2009.

The different data elements are partly defined in terms of the reporting guidelines while material types etc. are designated with reference to the relevant bibliographic standards specified within the Danish Cataloguing Rules¹⁵.

In addition to the more traditional statistics, information about visits to Danish public library web sites including OPACs and self-service library applications is also collected by the Ministry of Culture. The data collection is based on a widespread set of performance indicators, which individual libraries may adapt to local requirements.

Third party electronic content providers are strongly encouraged to provide *COUNTER*¹⁶ type usage statistics to Danish research libraries. The research libraries utilize relevant COUNTER statistics as background for the information that libraries submit to the national statistics.

Another significant development is that established types of data elements are supplemented by ad hoc information intended to gauge current trends and special activities within the public library sector. These special data elements, called *Library Barometer*¹⁷, are not necessarily intended to reflect established ISO standard concepts or long time series, but may cast light on relevant temporary and qualitative aspects.

4.3 Distribution of Information in Denmark

Statistical information is published by the collecting agencies as part of comprehensive reports and in the shape of selected data tables.

The reports are published as PDF-files, which can be downloaded for free.

The data tables are presented as predefined spreadsheets (EXCEL format), e.g. 236 attribute values (columns) for each of the 97 communes (rows) in Denmark.

The basic statistical information from research libraries and public libraries are published as two separate groups of Excel spreadsheets.

The consolidated annual statistics data for public libraries in Denmark is represented by the following sets of spreadsheets¹⁸:

- A set of 19 related spreadsheets (i.e. Excel workbook) with consolidated¹⁹ information about loans, holdings, interlibrary loans etc.
- A single spreadsheet with all of the raw data elements collected from each of the public libraries including regional libraries. I.e. 237 columns with attribute values and 98 rows representing each of the public library systems.
- Single spreadsheets with definitions that describe each of the 237 different data fields referenced within the raw data field columns mentioned above.

The consolidated annual statistics data for research libraries in Denmark is published within three similar groups of spreadsheets representing different groups of academic libraries²⁰.

Users may also compare and download selected statistical values via an interactive web application, where Key Performance Indicators for a selected commune or library are compared to corresponding national average values or time series²¹.

Statistical information about visits to Danish library websites is presented by a general interactive website, which incidentally also provides similar information about libraries in Sweden²².

4.4 Technical Solutions in Denmark

Statistical information about libraries in Denmark is compiled from several sources:

Some of the information about staff levels etc. is input by library representatives via a proprietary system with interactive web forms. Information about branch locations etc. is transferred automatically from other information systems operated by the Danish Agency for Libraries and Media.

Most of the data from public libraries is extracted automatically from the local library systems. This includes detailed information about loans and renewals, accession, holdings, interlibrary loans and active borrowers.

Transaction data is only aggregated at the lowest level of detail. I.e. several transactions are counted together, if they are identical except for one attribute value. Transaction data is exchanged via sets of text files with comma separated values. Each file contains only one type of information.

The data files are produced by the local library systems and attached to regular e-mail messages, which are sent to the collecting agency. Staff at the collection agency receives the e-mail messages and upload the attached files to the central data application system, which utilize a general purpose relational database system.

The central data import function performs various types of data validation which for example flags quantitative values that differ significantly from last year's corresponding values.

Before final publication, the collected data is made available via a web application, where the contributing libraries can inspect the data, which they have submitted to the central system. It is not possible for the contributing libraries to update their data on-line to the central system. Instead, any problems are resolved through direct contact with the collecting agency staff.

The central database includes statistical information and transaction data collected from the libraries during the last ten years. The transaction data is maintained at the original lowest aggregation level.

The statistical data about usage of public library web sites is aggregated by a separate system operated by a third party service provider. The usage data is collected by means of special scripts, which are embedded within the monitored web pages.

Some of the statistical information is forwarded to the Danish Statistics agency for inclusion in an integrated statistical data bank.

4.5 Systems Development and Operation in Denmark

Most of the systems used to collect and present statistical information concerning libraries in Denmark are implemented, maintained and operated by the national collecting agencies.

Interactive web forms used to collect information from the libraries are modified by internal staff members who also adapt the central import functions to import data from external data sources.

All of the relevant library systems can automatically export data files with the statistical information and transaction data for the national library statistics. These export functions are maintained by the local systems vendors based on annual specifications provided by the central collecting agencies.

National Reporting System for Library Statistics in Sweden

New material types etc. are accommodated automatically, because the data export/import functions are designed to reference the *danMARC2*²³ bibliographic exchange format.

Published data tables are extracted from the central system and converted to print formats (pdf) or two-dimensional spreadsheet tables (Excel). The resulting spreadsheets and reports are distributed via the Agency web site.

Statistical information about visits to public library websites is collected and published by means of a standard system offered by an external service provider²⁴.

4.6 Current Plans in Denmark

Danish Statistics will take responsibility for collection and publication of the Danish library statistics from the beginning of 2011, i.e. starting with the 2010 statistics.

Danish Statistics is expected to handle import of existing file formats with statistical data from local library systems.

Only libraries that effectively make their resources available to the general public will be included in future statistics. This change is expected to reduce the number of included research libraries by almost 50%.

The concept of a Library Barometer with ad hoc statistics and qualitative information will be pursued further as a supplement to the traditional library statistics with long standardized time-series.

4.7 Conclusions from Denmark

Much of the statistical information is exported and imported automatically from other national and local library systems in order to reduce errors and manual work. The current level of detailed information would be practically impossible to collect by manual methods.

Large amount of transactional information is stored at the lowest level of aggregation in order to preserve information, which might become relevant to future analysis.

Statistical data elements are defined in terms of relevant bibliographic data elements that are supported by local library systems. This facilitates automatic recording of new material types etc.

Special ad hoc statistics and qualitative data are in some cases considered more relevant than long time-series with a limited set of standardized quantitative data.

5 Library Statistics in Finland

The *National Library of Finland*²⁵ hosted a meeting in Helsinki on 2. September 2010.

The following representatives from Finland participated in the meeting:

- Tommi Jauhiainen, the National Library of Finland
- Markku Laitinen, the National Library of Finland
- Antti-Pekka Seppänen, *Uusimaa Centre for Economic Development, Transport and the Environment*²⁶

5.1 Organization in Finland

For historical reasons the national statistics about research- and public libraries in Finland are collected, managed and published by two different systems.

National statistics about research libraries in Finland are compiled by the National Library of Finland, which has a long tradition of managing national statistics for the research libraries.

The Public library statistics database in Finland is owned by the *Ministry of Education and Culture*²⁷. Development of the public library statistics is managed by an informal group including representatives of the Ministry of Education and Culture, *Centers for Economic Development, Transport and the Environment* (ELY) and *Helsinki City Library - Central Library for Public Libraries*²⁸

Public library statistics are collected in cooperation with regional authorities (ELY) and the 342 municipalities. Municipalities (libraries) fill in data, which is checked by the corresponding regional ELY offices.

ELY publishes the consolidated information together with various publications based on public libraries statistical data.

Approximately 70 institutions with research libraries submit statistical information directly to the National Library of Finland. Data for departmental libraries is included in the consolidated data provided by each hosting institution.

Hospital libraries for staff and researchers are included in the research library statistics, but the national statistic does not include information about patient libraries at hospitals.

Statistical information about research libraries in Finland is published by the National Library of Finland.

5.2 Information Collected in Finland

Public libraries in Finland collect a limited set of statistical indicators about physical loans, holdings, acquisitions and staff-levels etc. The statistical indicators are primarily directed towards administrative aspects and are not directly based on prevalent international standards such as ISO 2789.

For the sake of continuity, Finnish public libraries are hesitant to change or abolish existing statistical data elements, but aims to preserve long national time-series.

National Reporting System for Library Statistics in Sweden

Public libraries in Finland collect information about visits to their respective web-sites. The national statistics does not include other statistical information about usage of electronic resources at public libraries.

The national statistics does not include transactional level information imported automatically from operational IT systems in the public library sector.

In Finland, approximately 70 Research library institutions including special libraries submit ca. 400 data elements that are primarily based on the ISO 2789 and the ISO 11620²⁹ standards together with a few elements of ISO/TR 28118³⁰. These data elements are used to calculate more than 20 key indicators.

It is not feasible for libraries at some research institutions to give all information to the national statistics consequent upon the fact, that the science library statistics originally was created on the basis of the needs of university libraries.

Consolidated statistical information about loans and holdings at contributing research libraries is imported via data files exported from relevant local IT systems. For technical reasons the information is aggregated at a higher level than the detailed transaction data exported from Danish public library systems.

The semantic definitions of the statistical data elements are described within comprehensive manuals that also act as user's guides to the data entry applications.

5.3 Distribution of Information in Finland

Statistical information about libraries in Finland is processed and published via two different systems; one for public- and another for research libraries.

Finnish Research library statistics are published via an interactive web application, where users can pick selection criteria from drop-down lists organized in a hierarchical fashion³¹. The resulting statistics can be printed on-line or downloaded as Excel spreadsheets.

The public library statistics are also published via an interactive web application with dynamic selection criteria, but the public library web application seems a bit more flexible. Designated statistics can be printed on-line or downloaded as Excel spreadsheets³².

5.4 Technical Solutions in Finland

Statistical information about public libraries in Finland is entered via a proprietary system with interactive web-forms. The data is validated by the central system and stored within a general purpose relational database. The system also includes facilities to track the responses and to extract data for the interactive web queries.

Statistical data from research libraries in Finland is handled differently than the information from public libraries: Although some of the research library data is entered manually by library staff via interactive web forms, much of the statistical information is exported automatically from local library systems and imported to the central system.

The prevalent research library systems can export designated statistical data elements via files that are fed directly into the database of national system for research library statistics. The data files are formatted with

National Reporting System for Library Statistics in Sweden

a common XML format specified by the National Library. This XML exchange format is very simple and flexible with single-value self-identified data elements.

Contributing libraries may freely choose to submit their statistical information via the web forms and/or via export files. Regardless of the input method, all data elements can be updated or supplemented on-line via an interactive web application.

Plans are underway, so that research libraries may for example submit files with usage statistics exported from the *FinELib Consortium*³³ and manually add usage statistics from other consortia. This feature is currently being tested with data exported from selected systems and is expected to be fully functional before the end of 2010. Initially it will be feasible to handle data files exported from the *Voyager* ILS and the *FinELib Halti*³⁴ resource management system.

This approach is different from the related Danish solution, where exported data files contain more detailed transactional data than the Finnish files. While research libraries in Finland can each submit a few hundred statistical data elements via files and/or web forms, public libraries in Denmark each submit an average of over 7.000 elements via data files.

The persons in charge of statistics at contributing research libraries are recommended to check their respective statistical data by means of a special Excel tool. The National Library performs only spot checks on data submitted by the research libraries.

Research library statistical data is validated by means of special Excel macro programs before being accepted into the central statistics database.

Statistical information is published directly from the central systems via integrated interactive web applications. This accounts for the different user interfaces to research library data and public library data respectively, since the underlying integrated systems are different.

Since the initial version, the research library system has been enhanced with interactive facilities to define and modify organizational structures. This added flexibility permits the staff to reassign institutions and aggregate statistical information correspondingly.

5.5 Systems Development and Operation in Finland

The national system for research library statistics is called *KITT*. This is an integrated solution developed and maintained by a third party. Ongoing development and functional changes down to the addition of new statistical data fields therefore require additional contracts.

The national system for public library statistics is operated by a private company under contract with the Ministry of Education and Culture. The system was initially developed by the Ministry of Education and Culture together with the Provincial Administration, Helsinki City Library and the private contractor.

Minor changes to for example municipalities and other areas and to statistical data can be implemented by staff at the collecting agency (ELY), while bigger changes have to be ordered from the private company.

5.6 Current Plans in Finland

There are no plans to replace the national system for public library statistics, but some potential new features are under consideration in order to facilitate cross-sector regional analysis tasks etc.

- Common interface to a small part of the research library statistics and public library statistics.
- Geo-coding of relevant statistical information
- Data import functions using standard XML format
- Data export using a standard XML format

Work on the next version of the research library statistics system may begin in 2011, and a number of possible new features have been identified:

- Electronic resources handled on equal footing with physical resources
- Graphical visualization of statistical trends and time-series

The greater part of the e-resources of the Finnish libraries is acquired through the FinELib Consortium also known as the *National Electronic Library of Finland*. FinELib is planning an automatic system to collect usage statistics for the electronic resources. These usage statistics will be compiled according to the *COUNTER Codes of Practice*³⁵ and distributed via a web service compliant with the *ANSI/NISO Z39.93 Standardized Usage Statistics Harvesting Initiative (SUSHI)*³⁶ Protocol.

5.7 Conclusions from Finland

The collection of public library data is decentralized. Regional authorities compile statistical information from local public libraries and submit consolidated information to the national system for public library statistics.

The national system for public library statistics is separate and completely different from the national system for research library statistics, which is operated by the National Library of Finland.

Research libraries appreciate the option to submit information via data files exported from their local library systems.

Interactive web applications permit end-users to select relevant statistical information that is dynamically extracted directly from the respective integrated system databases. This is a useful alternative to large numbers of predefined tables.

A number of informal general recommendations were noted:

- Statistical data should be maintained at the lowest practical aggregation level
 - In order to facilitate future reuse of the statistical information.
- Different national authorities should harmonize exchange formats for statistical information
 - So that library systems vendors would be more motivated to support bespoke exchange formats.
- Statistical data should be exported and imported directly from relevant operational library applications.

National Reporting System for Library Statistics in Sweden

- In order to avoid manual data entry and to facilitate harvesting of detailed transactional statistics.
- Statistical information should be organized within a so called Data Warehouse.
 - In order to facilitate analysis across sectors, regions, material types and users.
- The user interface should provide intuitive and flexible facilities to perform custom queries and present the results in the shape of tables, spreadsheets and graphs.

6 Library Statistics in Norway

The *National Library of Norway*³⁷ hosted a meeting in Oslo on 17. September 2010.

The following representatives from the National Library of Norway participated in the meeting:

- Bjørnar Gundersen
- Erlend Ra
- Tertit Knudsen
- Torill Redse

6.1 Organization in Norway

The National Library of Norway is responsible for library statistics in Norway including public libraries, school libraries, research libraries and patient libraries etc. This responsibility together with relevant staff and other related resources was recently taken over from the *Norwegian Archive, Library and Museum Authority*³⁸.

The National Library coordinates the annual collection of statistical information from all participating libraries, but primarily communicates with representatives of research libraries, county libraries and state funded libraries. There are 323 research libraries in Norway plus 18 county libraries and 40 state funded libraries; i.e. a total of 381 libraries in addition to the public municipal libraries and school libraries.

Representatives of the 430 municipalities submit statistical data regarding public libraries and school libraries to their respective regional counties, which in turn forwards the consolidated information to the National Library. The municipalities in Norway have a total of 784 main- and branch public libraries.

Approximately 35% of the municipal public libraries are combined with one or more local school libraries, but only 10% of the school libraries are combined with a public library. I.e. 90% of the school libraries are independent entities. All loans from a combined library are statistically attributed to the corresponding public library part, while other activities and resources are divided between the public library- and the school library parts.

Norway has approximately 3.000 primary schools. Approximately 2.500 primary schools participated in the 2009 statistical survey. 150 of these reported that they did not operate a school library of their own.

Libraries for hospital staff etc. are counted among the research libraries while hospital libraries serving the patients are included among the state funded libraries.

Up until the statistics for 2009, the Norwegian Archive, Library and Museum Authority has compiled the statistical data information from all of the contributing libraries and has published a consolidated report with statistical information about archives, libraries and museums in Norway.

Data is also forwarded to *Statistics Norway*³⁹ for inclusion in the general national statistical databases.

6.2 Information Collected in Norway

Different types of libraries in Norway are requested to submit slightly different sets of statistical indicators to the national statistic.

The statistical indicators are primarily based on the ISO 2789 recommendations with some local modifications.

The semantic definitions of the statistical data elements collected are specified in terms of the corresponding data entry guidelines.

No detailed transactional datasets are collected for the national statistics.

The consolidated statistical indicators include general information about holdings and usage of electronic resources, but *COUNTER* type data elements are not yet collected for the national statistics.

6.3 Distribution of Information in Norway

General statistical information about libraries in Norway is published by the Norwegian Archive, Library and Museum Authority in the shape of predefined reports⁴⁰. The Norwegian Archive, Library and Museum Authority also publish more comprehensive information about research library statistics⁴¹. This information includes special reports and Excel spreadsheets.

Statistics Norway publishes statistical information about libraries in the form of predefined tables and reports. Some of the statistical data is also available via an interactive web application, where users can specify dynamic selection criteria. The resulting data can be presented as tables, graphs and Excel spreadsheets.

The statistical information about libraries in Norway may be obtained from the following sources:

- Statistical tables for research libraries⁴² in Norway
- Public libraries within the municipal statistics⁴³ for Norway
- Libraries within the culture statistics⁴⁴ for Norway

6.4 Technical Solutions in Norway

National library statistics in Norway are handled by an integrated system that covers collection of statistical information, but not publication of the statistical data.

The system includes facilities to specify the structure and content of interactive web forms used to collect statistical information from the libraries. The system also includes functions to manage collection tasks, track responses, validate data and export selected information to Excel spreadsheets, PDF-files and dynamic web pages.

The system utilizes a general relational database system. It is implemented by means of the Java programming language and other standard tools, which are supported by different system environments including *Windows* and *UNIX* (e.g. *SOLARIS*, *LINUX*) operating systems.

6.5 Systems Development and Operation in Norway

The national IT system for library statistics in Norway was developed in 2005 for the Norwegian Archive, Library and Museum Authority (ABM) with assistance from a number of external partners working under contract to ABM.

The system is designed with a *Service Oriented Architecture* (SOA) in order to facilitate systems maintenance and future integration with new services provided by ABM or third parties.

ABM owns the intellectual property rights to the proprietary code and makes it available under the terms of the *GNU General Public License*⁴⁵.

Although responsibility for the library statistics has been transferred to the National Library of Norway, ABM is still in charge of systems operation and maintenance.

ABM continues to handle national statistics for other types of cultural institutions including museums and archives with the same system.

6.6 Current Plans in Norway

There are no current plans to make any major changes to the existing national system for library statistics in Norway, but a number of new facilities are under consideration:

A web-based system for publication of statistical information would ease end-user access to the collected data. The system should cater both for advanced users, who require data for further analysis; as well as end-users, that prefer to use predefined tables and indicators.

Authorities wish to combine the library statistics with other economic-, demographic and geographical information in order to facilitate more holistic types of analysis.

The collecting agency also wishes to import statistical information via data files exported from local library applications. There are only a limited number of relevant library system vendors in Norway and they are generally willing to adopt common technical standards. It is therefore feasible to achieve widespread support for common technical guidelines concerning the exchange of statistical data between local library systems and the National Library.

6.7 Conclusions from Norway

It is helpful to perform automatic syntax checks and logical semantic controls during the data entry. But the interactive data collection forms and error messages must be carefully designed to avoid frustrating the users with overly zealous or confusing error messages.

It is important to provide self-service tools, so that staff members can customize data collection forms; specify new data elements and adjust aggregation hierarchies etc.

It is most practical to perform many types of ad hoc data validation and analysis by means of Excel spreadsheets.

It is desirable to export statistical data directly from relevant local library systems. And due to the nature of the library systems market in Norway, it is also feasible to do so in the future.

7 Library Statistics in Sweden

The Swedish Arts Council represented by Cecilia Ranemo hosted a meeting in Stockholm on 3. September 2010.

7.1 Organization in Sweden

The Swedish Arts Council is a government authority with responsibility for all national library statistics in Sweden – among other things.

The Swedish Arts Council collects annual statistics directly from different types of relevant libraries in Sweden including 40 research libraries, 37 special libraries, 80 hospital libraries and 290 municipal libraries.

The Swedish Arts Council contacts all of the designated libraries and collects the statistical data via web questionnaires, which are answered by local library staff.

Prior to each annual collection campaign, some library systems vendors are informed about any statistical indicators that are added or dropped from the current collection - relative to the preceding year. The system vendors are encouraged to modify their applications, so that their locally installed systems can produce the required statistical information.

Just as in other countries, school libraries in Sweden represent a particular challenge to the national library statistics: Out of more than 6.000 public school units, approximately 4.000 have access to some sort of school library. Half of the municipal libraries (ca. 500) are combined with school libraries.

There is no central registry of school libraries. The Swedish Arts Council periodically contacts more than 6.000 school units directly and asks them to submit statistical information about their respective school libraries. School libraries statistics were collected for 2008 and previously for 2002.

The Arts Council tracks the questionnaire replies from relevant libraries and validates the data. The contributing libraries are for example contacted directly, if the value of certain statistical indicators differ significantly from the previous year's values.

Data from the questionnaires is compiled by the Arts Council which publishes the consolidated national statistics in the form of printable reports and spreadsheets.

Some statistical data is also forwarded to *Statistics Sweden*⁴⁶, which publishes library statistics via an interactive web application with dynamic selection criteria.

7.2 Information Collected in Sweden

The Swedish Arts Council collects approximately 150 statistical indicators from each of the contributing public libraries. The statistical indicators are generally based on the ISO 2789 International Library Statistics, but not all of the recommended data elements are collected in Sweden.

Different types of libraries submit slightly different sets of statistical indicators. Historically, some of the statistical indicators have been added or dropped from year to year. Only some of the statistical indicators are therefore represented by longer time-series.

National Reporting System for Library Statistics in Sweden

Information about the annual statistics is documented within a set of formal declarations that include some of the definitions of statistical indicators collected from each type of library.

Data is collected via manual data entry through interactive web forms. Many of the local library systems can produce reports containing some of the required statistical information automatically, but no statistical information is exported directly to the central statistical databases via machine readable files. Because of this, transactional type data is not collected for the national statistics; only consolidated statistical indicators.

COUNTER type information is offered by some providers of electronic resources. This information is used by some libraries to calculate their usage statistics, but COUNTER data is not harvested automatically and forwarded directly to the national statistics system.

7.3 Distribution of Information in Sweden

The Swedish Arts Council publishes a set of reports and statistical data about libraries in Sweden.

Besides the reports, most of the statistical data is published in the shape of tables representing different types of libraries. I.e. data for public libraries, research libraries and hospital libraries is published within separate sets of tables⁴⁷.

Time-series from 1995 to 2009 with selected statistical indicators for public libraries is also available via a web page with an interactive graphic visualization tool⁴⁸.

Statistics Sweden publishes data about public⁴⁹ libraries via an interactive web application, where users can dynamically select information designated by predefined sets of hierarchical attributes. E.g. Number of Loans designated by Municipality, Material Type and Year.

7.4 Technical Solutions in Sweden

Statistical data is entered by participating library staff via interactive web survey forms. The web surveys are defined and managed by a standard application formerly called *SPSS Dimensions*⁵⁰.

All of the collected information is compiled in the form of SPSS or sometimes Excel spreadsheets. Unlike the other countries in this survey, Sweden does not use a general relational database system to manage its national library statistics data.

7.5 Systems Development and Operation in Sweden

The Swedish Arts Council maintains and operates the technical systems used to collect, compile and publish Swedish library statistics.

The Arts Council updates the interactive web forms used to collect statistical data from libraries and performs all of the data management tasks including validation, consolidation and publication of data sets used by the public.

7.6 Current Plans in Sweden

The future organization of national library statistics in Sweden is currently being investigated. In the meantime, the Swedish Arts Council currently continues to handle national library statistics in collaboration with the National Library of Sweden.

There are no specific plans to develop the current technical solutions any further.

The Swedish Royal Library has established a national *Expert Group for Library Statistics*⁵¹. This group has been the driving force behind a number of initiatives to explore a better national reporting system for library activities in Sweden.

The general issues and objectives for the proposed new system are described in a subsequent report called *Description of development project for official library statistics of Sweden*⁵². A draft version of this report was published in the beginning of 2010.

Requirements to the new system include:

- Facilities to import files with data compiled from local applications.
- Facilities to manage varying types of statistical indicators together with corresponding metadata
- Interactive tools to analyze statistical data involving different types of libraries and media types.

7.7 Conclusions from Sweden

The Swedish Arts Council collect and publish the statistics from several thousand libraries in Sweden by means of a few standard software tools and use relatively few resources to do so.

The statistical information to be collected varies from year to year and between types of libraries. It is therefore necessary to modify the corresponding interactive web forms used to enter the data.

Spreadsheets are useful to publish and analyze selected two-dimensional data tables, but a dedicated database system is better suited to manage large volumes of multi-dimensional statistical indicators with corresponding attributes and metadata.

Spreadsheets are the most popular format to publish statistical data, but not all libraries in Sweden use compatible local spreadsheet applications like MS Excel. It could therefore be useful to provide some additional web facilities for interactive data analysis.

8 Lessons from existing National Solutions

The investigation of existing national reporting systems for libraries yielded a number of lessons and insights which are relevant to the new solution in Sweden.

8.1 Organization

In some of the countries the collection of national statistics about libraries is managed by a single national organization, e.g. the National Library.

Publication of the official national library statistics is in some cases handled by the national statistics agency, which receives the data from the respective collection agency.

Some countries have organized the responsibility for national statistics according to the types of libraries: Municipal public library statistics are for example in some cases handled by the same organization as school library statistics.

School libraries are probably the most numerous types of libraries, but represent a special challenge to the national library statistics:

Many school libraries are partly integrated with municipal public libraries and do not collect separate statistics about all of the school library resources. Many school libraries are also run by a small informal staff which finds it difficult to collect extensive local statistics.

Some of the countries delegate responsibility for municipal library statistics to regional organizations, but none of the countries have apparently delegated responsibility for all types of libraries according to geographic criteria.

The statistics about libraries are generally managed by relatively few people within the national collection agencies.

Staff at the national collection agencies must act as consultants, which can advise library representatives and systems vendors about the statistical data elements to be collected.

It is helpful to delegate most of the management and consulting tasks to regional representatives or organizational collectives. Central agency staff can then communicate primarily with the regional representatives who in turn can assist local library representatives.

Library systems should be adapted to produce as much as possible of the statistical information automatically, so that local library staff can just forward information provided by their local systems.

The central collection agency must provide timely information about upcoming collections and new statistical indicators, so that system vendors can adapt their relevant library systems accordingly.

The scope of libraries involved in national statistics varies between countries. Sweden for example attempts to include as many libraries as possible, while Denmark applies more restrictive criteria.

The management of passwords and survey responses must be automated in order to minimize costly and error prone manual procedures involving up to several thousand libraries.

8.2 Information Collected

Most of the statistical indicators collected are based on relevant ISO standards supplemented by local interpretations and additional proprietary data elements.

Statistics about use of electronic resources are in most cases based on the ANSI COUNTER type statistical indicators, but are not yet imported directly from the e-content provider services via so-called SUSHI web services.

Different types of statistical indicators are collected from different types of libraries, e.g. municipal libraries versus academic libraries.

Some central agencies aim to compile long, consistent time-series. But it is fairly common to drop some elements and introduce new ones from time to time. This finding is very significant from a technical point of view:

The national statistical indicators do **not** conform to a static logical database schema with a limited number of different records types and data elements.

Much of the statistical data is presently collected via interactive web-surveys.

Finland and Denmark also import statistical data via files that are exported directly from local library systems. This is much more efficient than manual data entry. The success of this approach is dependent on close cooperation with library systems vendors in order to adapt the local systems to export the relevant data.

The file formats used to import statistical data includes proprietary text files, XML documents and standard spreadsheets (Excel).

Denmark collects large amounts of transaction level data exported from library systems, but other countries are also interested in collecting relevant data at the lowest possible level of aggregation – as opposed to summary statistics.

8.3 Distribution of Information

The national library statistics are generally published in the shape of standard printable reports as well as standard spreadsheet files that can be downloaded for local analysis and processing.

Most counties provide some interactive facilities to present different subsets of statistical data based on various selection criteria.

There is a general requirement for user-friendly facilities to present statistical information in the shape of graphs and diagrams.

Denmark provides special ad hoc statistics as a supplement to the more general standard statistical indicators. This approach has been very well received by the users.

8.4 Technical Solutions

All of the countries except Sweden use proprietary applications to collect-, manage and present the statistical data from libraries.

The systems provide different levels of flexibility to handle new types of statistical indicators and organizational changes, but adaptability is a major issue.

There is a strong requirement for self-service facilities, so that staff at the collecting agencies can adapt the systems to handle changing survey questions, organizational changes and new types of statistical indicators etc.

Some of the systems can import data via files that are exchanged via e-mail and uploaded to the respective central applications.

One of the systems in Finland allows libraries to upload some or all of the data elements via files with a general, extensible XML format and to edit the data via an interactive application. This approach seems to work very well.

Most of the systems provide some interactive facilities to search and present statistical data via interactive web applications. These applications exhibit different levels of user-friendliness and the graphical facilities are generally somewhat limited.

8.5 Systems Development and Operation

Most of the existing systems were developed in close cooperation with the collecting agencies but are now maintained and operated by separate IT organizations.

Operation and simple maintenance of the systems appears satisfactory, but ongoing development is a major issue. The challenges are not so much technical as organizational, when relatively minor changes may require costly and time consuming procurements based on binding formal specifications.

The current solution in Sweden reduces these issues by using general standard systems with built in customization facilities that can be adapted by in-house staff.

8.6 Current Plans

Most of the countries wish to import statistical data and relevant transaction level data directly from system generated files as a supplement to manual data entry.

The countries wish to combine the statistical library data with information from other sources including demographic- and finance data.

There is a particular requirement to link the statistical information with geographic criteria in order to facilitate the analysis of regional trends.

There is a general desire for better self-service facilities so that collecting staff can better adapt the solutions to handle emerging types of electronic materials, new statistical indicators and ad hoc analysis.

9 General System Requirements

Information and experiences learned from the analysis of existing national systems for library statistics has been used to identify general non-functional requirements to the new Swedish system.

The most important non-functional system requirements are identified within the following subsections.

9.1 Evolving Data Models

The statistical information collected from libraries on a national level does **not** conform to a uniform, static data model:

- A few statistical indicators are typically dropped or added to the surveys from year to year
- Different types of libraries supply different sets of information
- Emerging types of electronic materials appear within the usage statistics
- End-users wish to relate library statistics with data from other sources, e.g. demographic information, geographic attributes or financial data
- The popularity of the “Danish Library Barometer” illustrate the requirement to handle a rapidly evolving set of ad hoc statistical indicators in order analyze new emerging issues.

The system must therefore store and disseminate evolving sets of statistical data containing different types of data elements.

9.2 Diverse Data Sources

The national reporting system for libraries must import data from different sources:

- Response data collected via different survey services
- Statistical indicators prepared and validated via local systems
- Transaction data exported from local systems
- Usage data harvested from e-content service providers

As a result, the system must manage an increasing number of data files containing different types of data records – including metadata with information about the origin and structure of the datasets.

Initially, much of the statistical data collected from surveys and library systems etc. will probably be imported to the central system via batch files, which are produced by local systems and uploaded to the central system.

At some later time, it is likely, that some of the local systems will be modified to support special client functions, so that the local systems can submit data on-line to central server, i.e. without manual upload of batch files.

9.3 Different Dissemination Methods

The system must provide analytical data in different forms and with changeable contents.

- Evolving types of statistical information must be presented as performance indicators that can be reported as time-series or compared between different groupings

- Many users prefer to use a set of standard reports, tables and graphs based on a few relevant selection criteria
- The set of standard performance indicators, that are considered relevant, may change depending on available data sources and the issues, that are in focus at a certain time
- Some users need spreadsheet-like (e.g. Excel) tools to perform special analysis tasks with selected data

The system must provide intuitive access to standard reports and performance indicators. Spreadsheet programs are by far the most popular analytical tools.

The system must offer selected data in the shape of spreadsheets for download and must also provide some spreadsheet functions to users without suitable local spreadsheet programs.

9.4 Customization Facilities

Data types, input file formats and dissemination requirements to the national reporting system for libraries in Sweden are expected to evolve continuously.

It is therefore impractical to implement a static system where all of the functions are fixed from the start and where functional changes to data formats and user functions require extensive development activities.

The system must be customizable, so that professional staff can easily adapt the import and data management functions to handle evolving data types and file formats.

It must also be possible to specify interactive surveys, statistical indicators, tables, and reports via built-in user-functions.

9.5 Self-service Analysis Facilities

Professional staff may specify standard reports and common types of statistical indicators to be published via the central system.

In addition, some end-users may wish to perform more specialized types of analysis or to define special types of reports and graphics.

The system must provide self-service facilities so that professionals and advanced end-users can specify statistical indicators, reports and graphics with dynamic selection criteria.

It should also be possible to for the authors to share their specifications of statistical reports and analysis functions with other users via the central system.

9.6 Data Management

Initially, most of the data submitted to the national system for library statistics in Sweden will consist of relatively few statistical indicators and interactive surveys that produce a moderate amount of data. However, when libraries begin to submit large data files with selected transaction data, the total amount of data in the central system may grow by several orders of magnitude.

The main challenge is therefore to manage the processing and storage of a large number of different data sets with different types of data.

The system must include facilities to manage metadata and workflows in connection with import, documentation and storage of thousands of statistical datasets submitted from Libraries in Sweden.

9.7 System Deployment

There are several thousand libraries in Sweden, which potentially may submit-, update- and analyze information within the national system for library statistics.

The libraries have very different types of IT resources and use different and systems. Even common desktop applications such as spreadsheet programs used by libraries are not always compatible with widespread de facto standards.

Some of the major libraries may wish to implement local facilities in order to process data before upload to the central system, but it is in general impractical to install and maintain common software modules across all of the libraries in Sweden.

The national system must provide relevant facilities that are accessible as web applications and web services with standardized system interfaces and which do not require the libraries to deploy common system modules on their local systems.

10 Implementation Strategy

The proposed solution for the new national system for library statistics reflects the following insights:

The new system is **not** primarily required to process large numbers of uniform transactions in real time; this is unlike for example library circulation systems.

Instead, the main task is to manage thousands of datasets containing survey responses and statistical information exported from local systems. The format and content of these datasets may vary and evolve over time. While the volume of data may not be overwhelming, a key challenge is to consolidate semantically related information from different types of data sets– some of which are not yet known.

Custom made systems are usually most efficient for transaction oriented applications involving a large number of uniform transactions with relatively stable well-established business rules. But the national system for library statistics must manage different types of datasets and analytical tasks, which may evolve over time. The principal requirements to this type of system are flexibility and adaptability rather than performance.

10.1 Build on Standard System

For applications, which must handle different evolving types of data and analysis, it is often more efficient to adapt a general system with good self-service facilities (such as SharePoint). Rather than implementing a proprietary monolithic system from the ground up.

In order to attain the level of flexibility and adaptability necessary handle the different types of statistical data imported from a range of external systems, it is therefore proposed to base the new national solution on MS *SharePoint 2010*⁵³, which is a widespread general system that can be integrated with other external systems.

Flexible self-service facilities allow qualified staff to quickly adapt the solution to emerging new requirements – without going through a costly and time-consuming development effort.

In the beginning, a significant amount of systems design and development work may be required in order to implement the initial prototypes and to establish a sound foundation for future development. But the proposed strategy will probably require less time and resources than traditional systems development; especially when the cost of ongoing development and maintenance activities is taken into account.

Additional functionality can be added later via designated system parameters and customization of pre-existing parts of the solution.

10.2 Deploy Central Shared Services

SharePoint is deployed as a central service, which is accessible via standard web browsers over the Internet. It is therefore **not** necessary to install special software on local PCs and other systems etc. in order for the users to utilize the functions offered by the proposed solution.

Users without compatible local installations of MS Office may utilize relevant MS Office functions via the central SharePoint services. The central services can also interact directly with distributed local databases and other sources.

This strategy is more efficient and easy to maintain compared to alternatives that would require libraries to install local copies of common application modules.

System independence is an important general requirement to the new national system for library statistics, because users and libraries in Sweden employ a range of different local operation systems, library applications and desktop applications.

10.3 Develop Iterative and Incremental Solutions

The proposed solution can be implemented in stages and is designed to evolve over time, as staff and users make use of the system and identify new functional requirements.

Another benefit of customizing an existing standard system such as SharePoint is that it is fairly easy to implement prototypes to test different aspects of the intended solution.

A prototype utilizing the same standard products and design principles as the planned final system; may gradually evolve into a fully operational, scalable solution through a process of iterative development.

Implementation of the new system may therefore begin with the following steps leading to incremental functional prototypes:

1. Design the general logical structure of the central database. This structure forms the foundation of the rest of the system.
2. Specify general guidelines that describe data formats and procedures for data upload from surveys and local systems.
3. Establish a suitable test environment with relevant standard products and development tools, e.g. SharePoint with the SQL Server database system and Windows operating system with Internet Information Server
4. Implement an initial prototype to perform a limited beta test. A small set of survey response data is uploaded to the central system. Information is consolidated, edited and analyzed within the SharePoint framework system.
5. Implement another limited prototype, where a set of relevant statistical data is exported from a local application, uploaded to the central database and combined with information from some other relevant source.

11 Standard Products and Services

The proposed new national system for library statistics in Sweden is implemented on top of a few standard products and services.

It is proposed to utilize *Microsoft SharePoint Server 2010 for Internet Sites*⁵⁴ which is a general tool for data management, analysis and self-service solutions.

SharePoint integrates with the *MS SQL Server 2008 R2*⁵⁵ relational database management system and require the *MS Windows Server 2008 R2*⁵⁶ operating system.

These standard products are the current versions of widespread systems that are used and supported by a large number of independent organizations worldwide.

Commercial software licenses to the designated standard products may require a significant investment, but public authorities and educational institutions in Sweden may already have access to the products on more favorable terms.

All of the designated Microsoft software is usually freely available for test and evaluation purposes during a period of up to six months from the initial installation.

SharePoint provides a wide range of general tools to collect, manage, share, analyze and publish data via customized web applications.

Some users may prefer to utilize other specialized tools for certain tasks, e.g. interactive web surveys or web-site site design. This would not be a problem, because it is usually fairly simple to integrate different data sources and application with SharePoint based solutions.

12 Proposed Facilities and Services

This section describes general facilities and services offered by the proposed solution in relation to the main functions identified by the system architecture.

12.1 Survey Management

One of the common uses of SharePoint is for interactive web applications to collect, manage and share diverse sets of data.

A major advantage of using SharePoint for survey management is that SharePoint can also be used to integrate other related tasks; including user management and security, e-mail invitations, tracking of responses, data validation, analysis and feed-back.

There are also numerous alternative web survey tools on the market. Many of these tools are offered as free web services, where additional features are available for a price.

The current national solution to collect library statistics from libraries in Sweden utilizes a special survey management system called *SPSS Dimensions*⁵⁷.

Most standard survey management systems can export survey response data in the form of Excel spreadsheets, XML files and other widespread file formats, which can be imported into SharePoint.

12.2 Data Acquisition

Libraries may collect and prepare statistical information within proprietary local solutions. Libraries can for example consolidate data from local library applications, interactive surveys or *COUNTER*⁵⁸ type usage data; before the relevant statistical information is selected and exported to the national statistical database.

Statistical data can be imported from local systems and subsequently be edited within SharePoint. Alternatively, certain types of database systems support dynamic links to SharePoint facilities, so that local updates are automatically reflected within SharePoint.

SharePoint provides facilities to specify individual data sources and data transformation rules for importing statistical information from local systems and other data sources.

12.3 Data Management

The central database consolidates different types of shared statistical information originating from interactive surveys, local systems, and electronic content provider services (*SUSHI*⁵⁹) plus other sources.

Some types of relevant information may be difficult to represent by a static relational database schema. Unstructured-, dynamic- or proprietary datasets are examples of data sources that may be impractical to represent by a shared relational database.

SharePoint together with SQL Server includes metadata management facilities, so that information from different internal and external data sources can be combined to represent an integrated multi-dimensional model with dynamic semantic and hierarchical relationships between different data elements.

This feature can for example be used to analyze statistical information from libraries together with related geographical-, demographic- or financial information from other sources.

12.4 Information Dissemination

SharePoint offers facilities to present standard reports and “dashboards” with selected statistical indicators and graphs etc. that are updated automatically.

SharePoint also includes standard design templates and tools to manage interactive web sites to analyze and present information from different integrated data sources.

Within SharePoint web sites it is also possible to present information as dynamic Excel spreadsheets, which may be inspected on-line or downloaded to compatible local spreadsheet programs. Web users may access and manipulate Excel spreadsheets via a standard browser, i.e. without downloading the data to compatible local spreadsheet programs.

It is possible to control access to designated information within the central database, although secrecy is not a high priority, since national library statistics in Sweden are generally assumed to fall within the public domain.

13 Prototype Examples

An informal prototype has been developed to demonstrate some of the important features of the proposed solution.

- The most challenging requirement is to analyze dissimilar sets of statistical data together with geographical- and other types of information from different external sources.
- Another significant requirement is to support ad hoc analysis and to present the results in the shape of tables or graphs.

In order to illustrate, how these requirements can be tackled by the proposed solution, different types of relevant statistical- and geographical information from existing sources have been imported and linked automatically within an integrated multi-dimensional logical model.

(The prototype examples are only intended as a **technical** test. The data used may for example not be statistically representative and the data elements may not be comparable.)

The following example combines two different sets of statistical data about municipal libraries⁶⁰ and school libraries⁶¹. This information is published within separate sets of tables with different structures and types of statistical measurements.

Information about municipal libraries is linked with information about corresponding school libraries via geographical information about Swedish municipalities and counties, which is published by another government agency.

The different information sources are integrated within a multi dimensional logical data model that can be used for interactive analysis and presentation.

The following table shows the combined information about municipal libraries and school libraries aggregated by municipalities and counties.

It is possible to alter fields, selection-, sorting- and drill-down criteria interactively as illustrated by the drill-down to display data for individual municipalities within Blekinge county.

Rækkenavne	Municipal library locations	Schools with own libraries
Blekinge län	37	51
Karlshamn	5	13
Karlskrona	11	16
Olofström	3	8
Ronneby	14	7
Sölvesborg	4	7
Dalarnas län	113	113
Gotlands län	15	25
Gävleborgs län	101	78
Hallands län	48	78
Jämtlands län	67	24

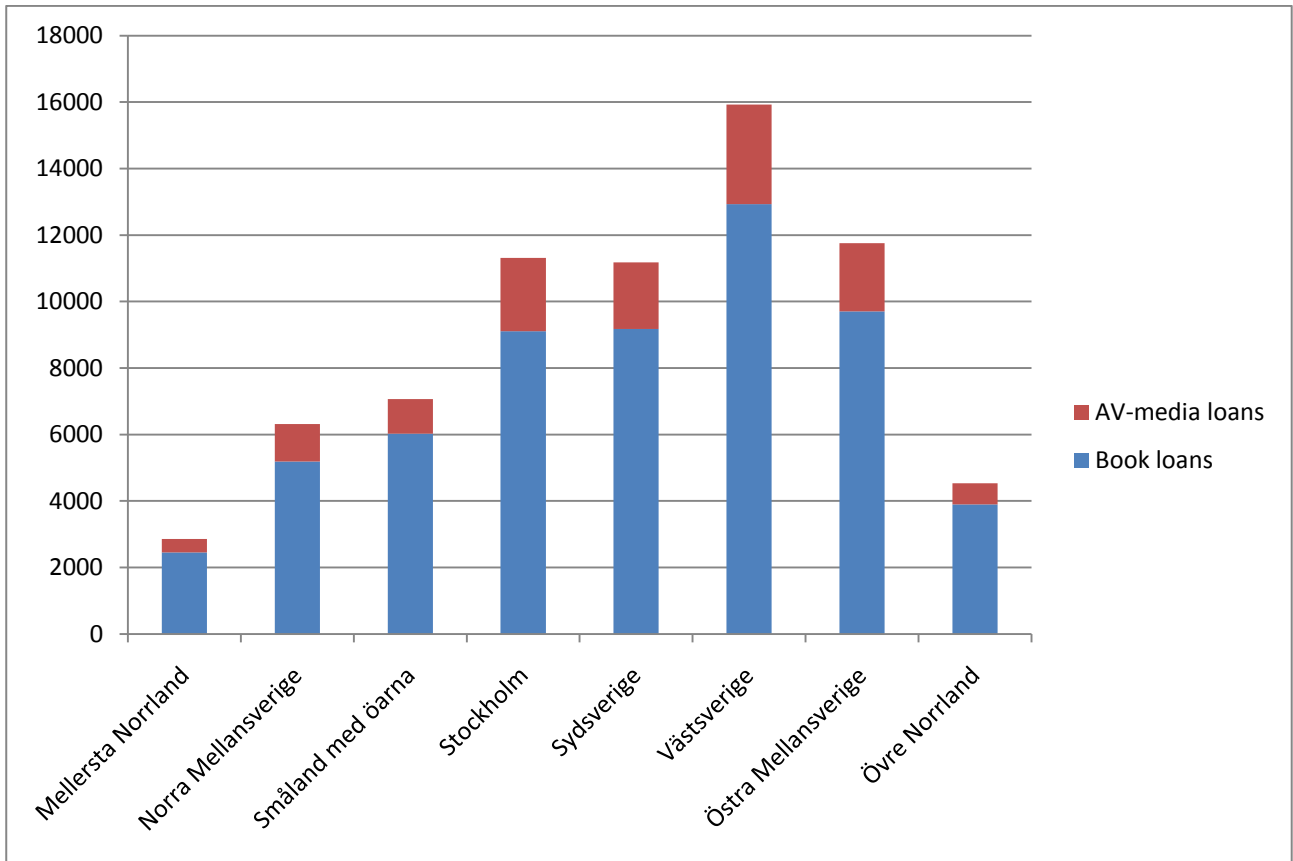
National Reporting System for Library Statistics in Sweden

Jönköpings län	133	123
Kalmar län	96	74
Kronobergs län	55	57
Norrbottnens län	159	56
Skåne län	303	389
Stockholms län	236	482
Södermanlands län	52	86
Uppsala län	60	81
Värmlands län	52	92
Västerbottens län	128	76
Västernorrlands län	59	59
Västmanlands län	58	77
Västra Götalands län	351	461
Örebro län	85	87
Östergötlands län	63	153
Hovedtotal	2271	2722

Analytical data and other results can automatically be presented as tables, reports, spreadsheets or interactive diagrams.

The following diagram illustrates existing published data sources that are aggregated automatically according to an external hierarchical geographical classification system called *NUTS*⁶², no pun intended.

National Reporting System for Library Statistics in Sweden



The current informal prototype utilizes some of the same standard tools for data management and multi-dimensional analysis products as the proposed solution.

This prototype use data from the national Swedish library statistics and other relevant external sources in order to make the functional test as realistic as possible.

The amount of data, which is available as part of the published Swedish national statistics for libraries, is too small to represent any sort of challenge to the capacity of the prototype system. But other tests using the same tools indicate, that the proposed solution can easily scale to handle datasets several orders of magnitude bigger.

14 Notes

- ¹ National Library of Sweden Expert Group on Library Statistics.
<http://www.kb.se/bibliotek/referensgrupper/expertgrupp-statistik/>
- ² National Library of Sweden. <http://www.kb.se/english/>
- ³ Swedish Arts Council. <http://kulturradet.se/en/In-English/>
- ⁴ Lund University Libraries. <http://www.lu.se/bibliotek>
- ⁵ Functional Requirements to Swedish Library Reporting System.
<http://www.kulturradet.se/Documents/statistik/bibliotek/Functional%20Requirements%20to%20Swedish%20Library.pdf>
- ⁶ COUNTER Usage Statistics. <http://www.projectcounter.org/about.html>
- ⁷ SUSHI Protocol.
http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=2de0e3e04f3a7e32d45db8ee87574c3c8206ddcb
- ⁸ The Danish Agency for Libraries and Media. <http://www.bibliotekogmedier.dk/english/>
- ⁹ Aggregate data. http://en.wikipedia.org/wiki/Aggregate_data
- ¹⁰ Danish Ministry of Culture Library Statistics. <http://www.bibliotekogmedier.dk/biblioteksomraadet/statistik/>
- ¹¹ Danish Ministry of Education Library Statistics.
<http://www.undervisningsministeriet.dk/service/Statistik/Folkeskolen%20og%20frie%20skoler/Skoler/Skolebiblioteker.aspx>
- ¹² Statistics Denmark Library Statistics. <http://www.dst.dk/pukora/epub/upload/15197/headword/dk/98.pdf>
- ¹³ ISO 2789:2006. International Library Statistics. http://www.iso.org/iso/catalogue_detail.htm?csnumber=39181
- ¹⁴ Reporting Guidelines for Danish Library statistics.
<http://www.bibliotekogmedier.dk/biblioteksomraadet/statistik/statistikvejledninger/>
- ¹⁵ Danish Cataloguing Rules. <http://www.kat-format.dk/>
- ¹⁶ Counting Online Usage of NeTworked Electronic Resources. <http://www.projectcounter.org/index.html>
- ¹⁷ Danish Library Barometer. <http://www.bibliotekogmedier.dk/statistik/biblioteksstatistik/biblioteksbarometret/>
- ¹⁸ Danish Public Library Statistics data. <http://www.bibliotekogmedier.dk/statistik/biblioteksstatistik/folkebiblioteker/>
- ¹⁹ Data consolidation. <http://www.merriam-webster.com/dictionary/consolidating>
- ²⁰ Danish Research Library statistics data.
<http://www.bibliotekogmedier.dk/statistik/biblioteksstatistik/forskningsbiblioteker/>
- ²¹ Interactive Comparison of Danish Library Statistics. <http://sbs.bs.dk/>
- ²² Danish Library Website visitor statistics.
http://bib.kpiindex.dk/KPIBibliotekerne_2010/?uge=37&compareLibrary=true
- ²³ danMARC2 bibliographic exchange format. <http://www.kat-format.dk/danMARC2/default.html>
- ²⁴ KPI Index Web statistics. <http://www.kpiindex.dk/index.php>
- ²⁵ The National Library of Finland. <http://www.nationallibrary.fi/libraries/statistics.html>
- ²⁶ Centres for Economic Development, Transport and the Environment in Finland. <http://www.ely-keskus.fi/fi/Sivut/inenglish.aspx>
- ²⁷ Finland Ministry of Education and Culture. <http://www.minedu.fi/OPM/Kirjastot/?lang=en>
- ²⁸ Helsinki City Library. <http://www.lib.hel.fi/en-GB/>
- ²⁹ ISO 11620:2008 Library performance indicators.
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37853
- ³⁰ ISO 28118:2009 Performance indicators for national libraries.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=44512
- ³¹ Finnish Research Libraries Statistics.
http://yhteistilasto.lib.helsinki.fi/language.do?action=change&choose_language=3
- ³² Finnish Public Libraries Statistics. <http://tilastot.kirjastot.fi/en-GB/>
- ³³ Finnish National Electronic Library. <http://www.nationallibrary.fi/libraries/finelib/finelibconsortium.html>
- ³⁴ HALTI Resource Management System.
http://igelu.org/files/webfm/public/documents/conference2007/18b_pennanen.pdf
- ³⁵ COUNTER COdes of Practice. http://www.projectcounter.org/code_practice.html

- ³⁶ ANSI/NISO Z39.93 Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol.
http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=2de0e3e04f3a7e32d45db8ee87574c3c8206ddcb
- ³⁷ The National Library of Norway. <http://www.nb.no/english>
- ³⁸ Norwegian Archive, Library and Museum Authority. http://www.abm-utvikling.no/?set_language=en
- ³⁹ Statistics Norway. <http://www.ssb.no/english/>
- ⁴⁰ Norwegian Library Statistics. <http://www.abm-utvikling.no/bibliotek/statistikk-for-bibliotek>
- ⁴¹ Norwegian Research Library statistics. <http://www.abm-utvikling.no/bibliotek/statistikk-for-bibliotek/statistikk-for-fag-og-forskningsbibliotek>
- ⁴² Research libraris at Statistics Norway. http://www.ssb.no/english/subjects/07/01/40/ffbibl_en/
- ⁴³ Norwegian Public library statistics. http://www.ssb.no/english/subjects/00/00/20/kostra_en/
- ⁴⁴ Libraries in Norwegian culture statistics. <http://www.ssb.no/english/subjects/07/>
- ⁴⁵ GNU General Public License. <http://www.gnu.org/licenses/gpl.html>
- ⁴⁶ Statistics Sweden. http://www.scb.se/default_2154.aspx
- ⁴⁷ Library Statistics from Swedish Arts Council. http://www.kulturradet.se/en/publikationer_/Bibliotek-2009/
- ⁴⁸ Visualization of Swedish Public Library time-series data. [http://statistik.kulturradet.se/folkbibliotek/199 from 1995 to 2009 5 2009.htm](http://statistik.kulturradet.se/folkbibliotek/199%20from%201995%20to%2009%205%202009.htm)
- ⁴⁹ Statistics Sweden Public Libraries data. http://www.scb.se/Pages/ProductTables_10221.aspx
- ⁵⁰ IBM SPSS Data Collection. <http://www.spss.com/software/data-collection/>
- ⁵¹ The Swedish Royal Library Expert Group for Library Statistics.
<http://www.kb.se/bibliotek/referensgrupper/expertgrupp-statistik/>
- ⁵² Development project for official library statistics of Sweden.
http://www.kb.se/dokument/Bibliotek/referensgrupper/biblioteksstatistik/Biblioteksstatistikpunktse_100225.pdf
- ⁵³ MS SharePoint 2010 Sverige. <http://sharepoint.microsoft.com/sv-se/Sidor/default.aspx>
- ⁵⁴ MS SharePoint Server 2010 for Internet Sites. <http://sharepoint.microsoft.com/en-us/internetsites/products/Pages/SharePoint.aspx?Product=SharePoint>
- ⁵⁵ MS SQL Server 2008 R2. <http://www.microsoft.com/sqlserver/2008/en/us/R2-self-service-BI.aspx>
- ⁵⁶ MS Windows Web Server 2008 R2. <http://www.microsoft.com/windowsserver2008/en/us/2008-web.aspx>
- ⁵⁷ IBM SPSS Data Collection. <http://www.spss.com/software/data-collection/>
- ⁵⁸ COUNTER Usage Statistics. <http://www.projectcounter.org/about.html>
- ⁵⁹ The Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol.
http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=2de0e3e04f3a7e32d45db8ee87574c3c8206ddcb
- ⁶⁰ Sweden Municipal library tables 2009.
http://www.kulturradet.se/Documents/statistik/bibliotek/Biblioteksstatistik_2009/Folkbibliotek_tabeller%20efter%20kommun_2009.xlsx
- ⁶¹ Sweden School library tables 2008.
http://www.kulturradet.se/Documents/publikationer/2009/Skolbibliotek_bilaga_4_delresultat_kommuner_lan_%c3%a4ndrad090821.xls
- ⁶² NUTS - Nomenclature of territorial units for statistics.
http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction