

FUNCTIONAL REQUIREMENTS TO SWEDISH LIBRARY REPORTING SYSTEM

INNEHÅLL

1.	FÖRORD PÅ SVENSKA	5
2.	INTRODUCTION	6
2.1.	Purpose of this study	6
2.2.	Background of the study	6
2.2.1.	Project Proposal	6
2.2.2.	Steering Group	7
2.2.3.	Objectives	7
2.2.4.	Standards	7
2.2.5.	Library Statistics	8
2.2.6.	Methodology	8
2.2.7.	Library Representatives	9
2.3.	System Scope	11
3.	USER REQUIREMENTS	12
3.1.	Assumptions and Conditions	12
3.2.	User Characteristics	13
3.3.	User Functions	14
4.	GENERAL SYSTEM DESCRIPTION	17
4.1.	System Context	17
4.2.	Major System Features	20
4.3.	Major System Constraints	20
4.4.	Operational Scenarios	21
5.	FUNCTIONAL REQUIREMENTS	22
5.1.	Manage Surveys	22
5.1.1.	Specify Survey Content	22
5.1.2.	Customize Survey Format	22
5.1.3.	Manage Survey Invitations	23
5.1.4.	Manage Survey Responses	23
5.2.	Collect Surveys	24
5.3.	Manage Metadata	24
5.4.	Analyze Data	25
5.5.	Upload Transaction Data	25
5.5.1.	Specify Transaction Datasets	25
5.5.2.	Upload Transaction Datasets	26
5.5.3.	Upload Data Clients	26
5.6.	Harvest Usage Data	26
5.7.	Communication Interface Requirements	26
6.	NONFUNCTIONAL REQUIREMENTS	28
7.	SUMMARY AND RECOMMENDATIONS	29
7.1.	User Requirements	29
7.2.	System Overview	29
7.3.	Metadata dictionary	29
7.4.	Repository Database	30
7.5.	Survey Management	30

7.6.	Data Management	30
7.7.	Data Import	30
7.8.	Data Analysis and Export	30
7.9.	Next Steps	31

1. FÖRORD PÅ SVENSKA

Under 2009 uppdrog Kungliga bibliotekets expertgrupp för biblioteksstatistik IT konsulten Poul Henrik Jørgensen att genomföra en förstudie om hur en teknisk plattform för framtida svensk biblioteksstatistik borde konstrueras. Denna rapport är skriven av den samme och de slutsatser och förslag som ges är de som han funnit genom sitt arbete med studien. Arbetet har bekostats av Statens kulturråd och KB, Kungliga biblioteket.

Studien genomfördes via personliga intervjuer med nyckelpersoner vid svenska bibliotek samt via webbenkäter till de som inte kunde medverka vid personliga intervjuer. Studien är i sin form en inventering av vilka (tekniska) funktionella krav som borde ställas på ett framtida system som kan inhämta och sammanställa statistiken för verksamheten som de offentligt finansierade biblioteken i Sverige har.

Det finns ingen svensk översättning av rapporten. Förfrågningar om studien kan göras till Cecilia Ranemo på Kulturrådet, cecilia.ranemo@kulturradet.se.

2. INTRODUCTION

2.1. Purpose of this study

This report describes general requirements to a national system to streamline the collection, storage and delivery of statistical data related to library activities and services. The requirements were compiled as part of a feasibility study to investigate the circumstances for developing a common reporting system for libraries in Sweden.

This feasibility study was initiated by the Swedish National Council for Libraries¹ following a proposal put forward by the National Library of Sweden's Expert Group on Library Statistics². The study was funded jointly by the National Library of Sweden³ together with the Swedish Arts Council⁴.

The study was conducted by IT consultant by Poul Henrik Jørgensen⁵ who also prepared this report during August to November 2009.

An earlier draft of this report was presented to the Expert Group at a meeting in Stockholm on 19. November 2009. The present draft version was first presented at a meeting of the Swedish National Council for Libraries in Stockholm on 3. December 2009.

The intended audience for this report includes members of the National Council and the Expert Group plus others with an interest in the establishment of the common statistical reporting system for libraries in Sweden.

2.2. Background of the study

Libraries face mounting demands to document the efficiency and impact of library services. But in order to perform a quantitative assessment of library performance, it is necessary to compile accurate and timely measurements and performance indicators.

2.2.1. Project Proposal

The Expert Group for library statistics proposed to initiate a feasibility study to investigate the conditions for development of a tool to monitor library activities. The Project Proposal⁶ was prepared in November 2008 and the resulting Project Description was ready in June 2009⁷.

¹ Sverigebiblioteket. <http://www.kb.se/bibliotek/referensgrupper/sverigebiblioteket/>

² KB:s Expertgrupp för biblioteksstatistik. <http://kb.se/bibliotek/referensgrupper/expertgrupp-statistik/>

³ Kungl. biblioteket - Sveriges nationalbibliotek. <http://kb.se/om/>

⁴ Statens kulturråd. <http://www.kulturradet.se/sv/Om-kulturradet/>

⁵ Poul Henrik Jørgensen, Portia I/S. E.mail: PHJ(at)Portia.dk Phone: +45 49181908

⁶ Utveckling av en teknisk plattform för biblioteksstatistik. Förslag till förstudie 2008-11-13.

http://www.kb.se/dokument/Bibliotek/referensgrupper/biblioteksstatistik/forstudie_teknisk_plattform.pdf

⁷ Förstudie för teknisk plattform för biblioteksstatistik. Statens Kulturråd Dnr Adm 2009/94

2.2.2. Steering Group

The steering group for the project consists of Cecilia Ranemo⁸ and Erik Peurell⁹ from the Swedish Arts Council, Christine Lindmark¹⁰ from the National Library of Sweden and Tore Torngren¹¹ from Lund University Libraries¹².

2.2.3. Objectives

The feasibility study is expected to identify general requirements to a system to collect, manage and extract statistical information about library activities and services. The system must serve the needs of different types of publicly funded libraries including public-, school-, hospital- and academic libraries.

The system must also serve as flexible repository for statistical measurements and transaction data collected by libraries, so that analysts can cross tabulate different types of information and compare trends in relation to time, geographical regions and other dimensions.

2.2.4. Standards

Although this study is not concerned with deciding which measurements or transactions to collect, it is helpful to reflect on the most relevant data standards that are currently in use:

- The ISO 2789 standard for *International library statistics*¹³ defines some general quantitative measurements of library resources and usage.
- ISO 11620 *Library performance indicators*¹⁴ describe measurements for libraries in general, while the eponymous ISO/TR 28118¹⁵ is limited to *Performance Indicators for National Libraries*.
- Some providers of electronic content calculate usage data according to the *COUNTER Codes of Practice*¹⁶, which has attracted fairly widespread support among third party Electronic Content Providers and libraries.

These standards define a useful framework and methodology for assessment of library activities. But the increasing differentiation of library services coupled with new types of library objects may render some of the specified indicators obsolete or inadequate. It is therefore necessary to allow for additional types of measurements and performance indicators in the future.

⁸ Cecilia Ranemo. <http://www.kulturradet.se/sv/statistik/bibliotek/>

⁹ Erik Peurell. <http://www.kulturradet.se/sv/statistik/bibliotek/>

¹⁰ Christine Lindmark. <http://www.kb.se/bibliotek/statistik/>

¹¹ Tore Torngren. <http://www.lu.se/o.o.i.s?id=7130&task=listPerson&username=lub-tto>

¹² Lunds universitets bibliotek. <http://www.lub.lu.se/>

¹³ ISO 2789:2006 International library statistics.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=39181

¹⁴ ISO 11620:2008 Library performance indicators.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=37853

¹⁵ ISO/TR 28118:2009 Performance indicators for national libraries.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=44512

¹⁶ COUNTER Codes of Practice. http://www.projectcounter.org/code_practice.html

2.2.5. Library Statistics

The Swedish Arts Council is designated as the national authority for library statistics in Sweden¹⁷. The Arts council compiles information from public-, school- and hospital libraries plus certain other libraries¹⁸.

Sweden currently has 290 public municipal libraries with a total of 2097 lending locations.

Approximately 4.000 schoolunits have access to school library facilities, some of which are shared among several schools. 548 public municipal library units are integrated with school libraries.

The Royal Library produces annual statistics about academic libraries including University-, College and Research libraries plus Special libraries¹⁹.

Sweden has 40 University- and College libraries plus 37 special libraries including museums, research institutions and public authorities. The research libraries provide a total of 208 service locations.

2.2.6. Methodology

This study includes an inventory of functional user requirements plus descriptions of major system functions. The resulting specifications may be used as basis for a subsequent procurement process and systems development.

In order to leave room for different system architectures and implementation strategies, the specifications are mostly concerned with general requirements and avoid aspects, which are not necessitated by the user requirements.

In September 2009 the Swedish government proposed the establishment of a new independent authority with responsibility for analysis and evaluation of cultural activities from 2011²⁰. The future organizational framework for collection and management of library statistics is therefore unknown at the time of writing.

It might be awkward to specify detailed nonfunctional system requirements, before the future authority in charge of the system has been identified. Consequently these specifications attempts to stay away from special requirements that might unduly restrict the system design including the choice of implementation method and operating environment.

User requirements to the new system were collected through a series of meetings with users representing different types of libraries. Invitations were distributed to a list of 43

¹⁷ Statistik om Biblioteken. <http://www.kulturradet.se/sv/statistik/bibliotek/>

¹⁸ Statistiska centralbyrån Folkbiblioteksstatistik. http://www.scb.se/Pages/Product___10211.aspx

¹⁹ Kungl. biblioteket Statistik och kvalitet. <http://kb.se/bibliotek/statistik/>

²⁰ Kulturpropositionen. <http://www.kulturradet.se/Nyheter1/Kulturpropositionen-Tid-for-Kultur/>

library executives proposed by the steering committee. Thirteen representatives participated in the interview meetings conducted by Poul Henrik Jørgensen at the Arts Council during September and October 2009.

Participants were presented with a list of questions about general assumptions and functional user requirements. Designated library representatives, who were unable to participate in the interview meetings, were invited to participate in a web-survey²¹ with the same questions.

Participants were also asked to forward the invitation to other relevant user representatives. Fifteen additional library representatives participated in the web survey.

The user requirements were identified on the basis of the user interviews and the web survey supplemented by inputs from members of the project steering group.

2.2.7. Library Representatives

28 library representatives contributed directly to the requirements specification thus providing a total of 700 answers to the questions about assumptions and functional requirements to the new system.

The 28 contributors represent the following mix of library types:

- 11 academic libraries (red)
- 5 public libraries (cyan)
- 5 regional libraries (yellow)
- 4 hospital libraries (blue)
- 3 school libraries (green)

The geographic distribution of the represented libraries is indicated on the following map.

Pin colors refer to the library types listed above.

Click to access [Interactive map of participating libraries](#)

²¹ Användarnas krav på teknisk plattform för biblioteksstatistik.
<http://surveys.polladdy.com/s/1FC3BB2441779564/>

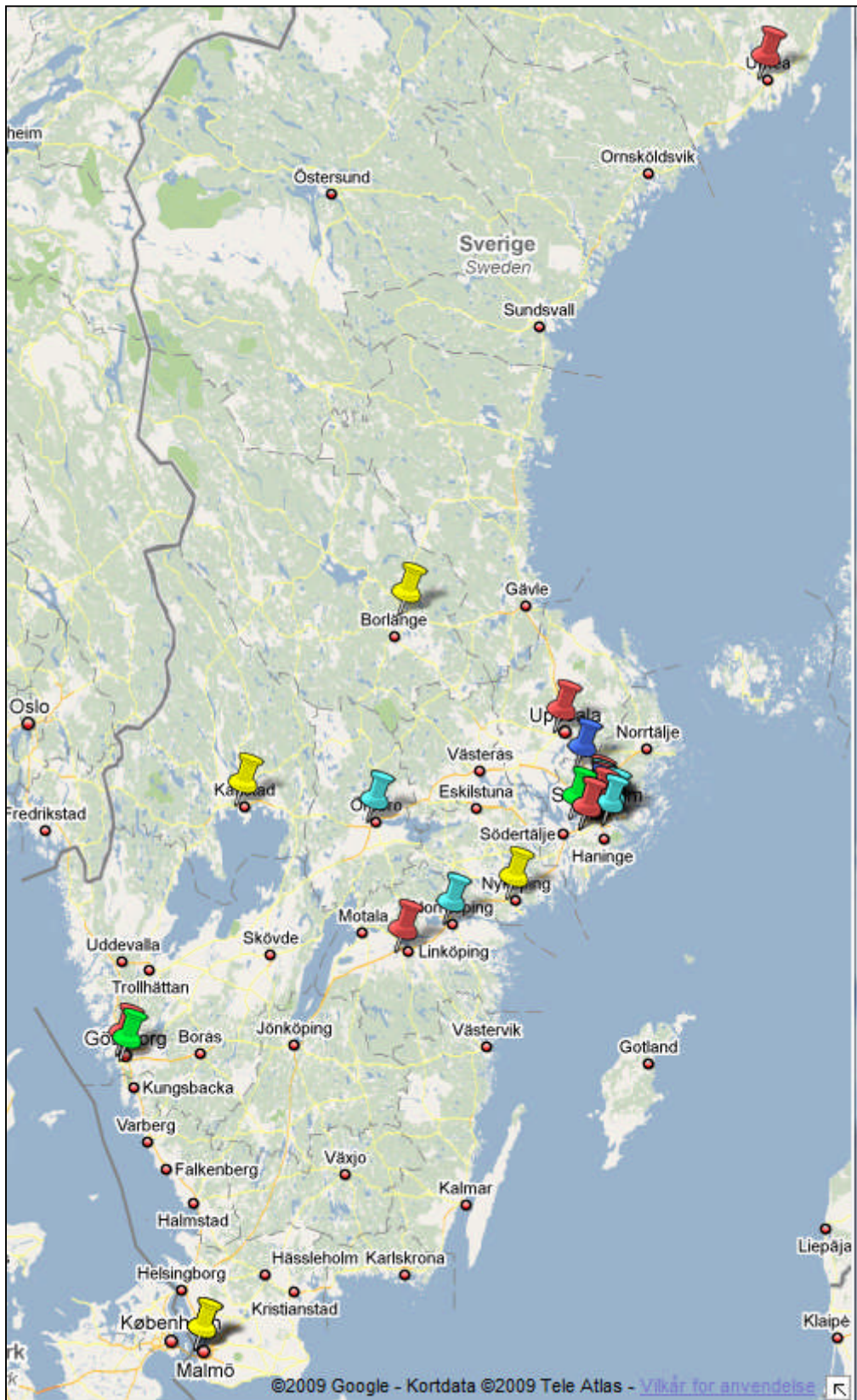


Fig. 1 Library User Representatives

2.3. System Scope

The primary goal of the specified system is to make it easier for publicly funded libraries in Sweden to monitor designated performance indicators and to access the impact of library activities.

The system will be used to manage the collection, dissemination and analysis of Swedish national library statistics in. But it must also support monitoring and assessment of local library activities based on specialized proprietary statistical measurements, transaction data and performance indicators.

Data measurements collected within the system must be usable for analysis tasks today and into the foreseeable future, long after the system itself may be replaced. One of the most important requirements is therefore to collect and maintain necessary metadata, which will facilitate analysis, integration, maintenance and conversion of the data.

3. USER REQUIREMENTS

3.1. Assumptions and Conditions

These specifications are based on a number of general assumptions, which have been evaluated by user representatives during interview meetings and as part of the aforementioned web survey²¹.

The user representatives were asked to pass verdict on a list of six general assumptions. The designated assumptions with tabulated answers are listed below. Possible responses were *Yes* (Y), *No* (N) and *Unknown* (?)

Assumption	Yes	No	?
a1 Libraries will use the collected information for local analysis and decision making	27	0	1
a2 Central authorities will collect shared statistical information from all libraries.	26	0	2
a3 Libraries will collect proprietary information about selected libraries and activities.	23	0	5
a4 The system must handle transaction datasets in addition to statistical data elements.	21	0	7
a5 Libraries have limited resources to adapt local systems.	24	2	2
a6 Libraries and other users will prefer to download selected data from the central system and to edit, analyze and present the data by means of standard spreadsheet programs and other local tools.	17	2	9

The following graph illustrates the responses to each of the six assumptions.

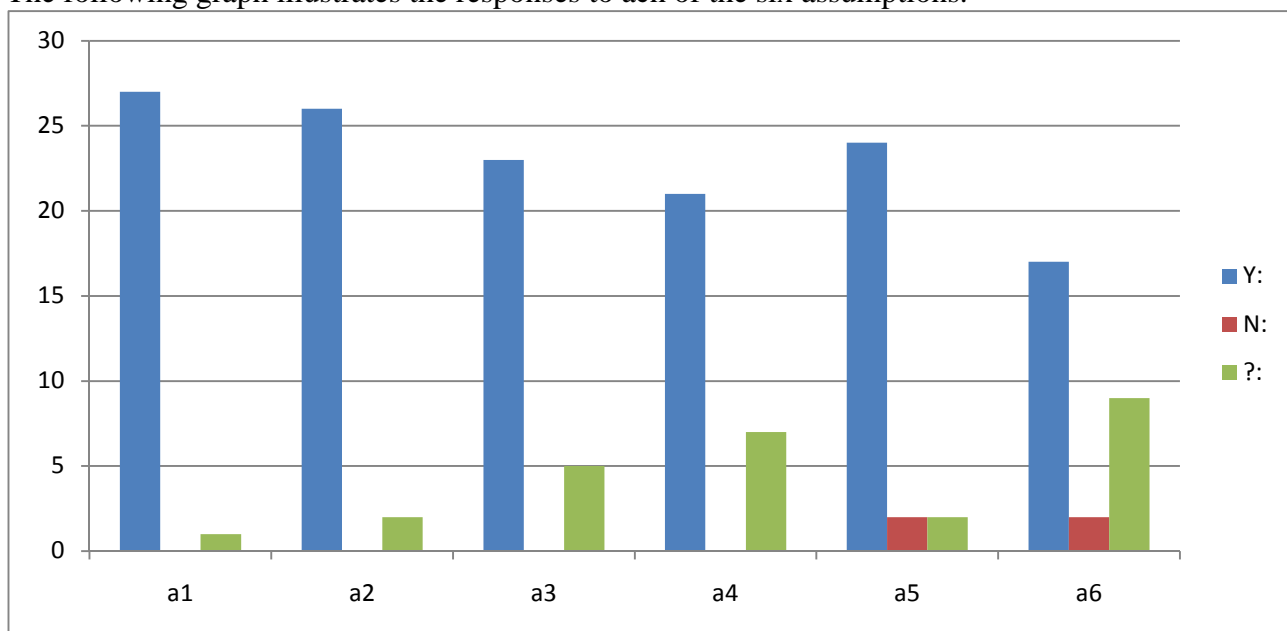


Fig. 2 Assumptions Evaluation

Evidently, respondents generally agree with the proposed assumptions. The only exception is assumption number six. I.e. 39% of the respondents are unsure about the need to download data to spreadsheets and other local analysis tools.

The most noteworthy result is the strong consensus about assumption number three: I.e. the system must also handle local proprietary types of statistical information and measurements in addition to shared data collected by national authorities. This requirement represents a significant technical challenge.

3.2. User Characteristics

The system must serve different types of users:

- **System managers:** Skilled IT professionals, which manage the system including Repository databases, central web applications and services.
 - System managers are designated as part of a central service agreement.
 - Prefer English language interface.
 - Estimated max. 10 active system managers during first year of operation.
 - The number of persons required to manage the system is dependent on how the system is implemented and operated.
 - Some comparable systems are managed by a small handful of staff.
- **Data Managers:** Trained library staff, who enters and updates analytical data owned by the library.
 - Each library needs a data manager to manage its data measurements.
 - Require Swedish language interface.
 - Estimated 8.000 active data managers during first year.
 - Assuming, that each of the approximately 5.000 libraries will designate 1-2 local staff persons to enter and update statistical data for the library.
- **Metadata managers:** Specialized library professionals, who specify metadata and repository structures.
 - A library needs a metadata manager, if the library wishes to specify local proprietary types of statistical measurements and transaction data.
 - Prefer Swedish interface, but English is acceptable.
 - Estimated max. 50 active metadata managers during first year.
 - Assuming less than 40 libraries plus the central authorities will allocate staff to manage local and shared metadata.
- **Survey managers:** Trained library professionals, who specify surveys, manage invitations and track responses.
 - A library needs a survey manager, if the library wishes to collect local surveys.
 - Prefer Swedish interface, but English is acceptable.
 - Estimated max. 200 active survey managers during first year.

- Assuming less than 200 libraries plus the central authorities will assign staff to manage local and shared surveys.
 - Potentially all libraries are likely to conduct their own user surveys sometime in the future.
- **Upload managers:** Trained IT professionals who manage data extracted from local applications and uploaded to the via the central upload web application.
 - Upload managers may be allocated as part of shared local service agreements.
 - Swedish interface is preferred but can use English.
 - Estimated max. 20 active upload managers during first year.
 - Assuming less than 20 libraries plus central authorities will assign staff to manage upload of datasets.
- **Survey Respondents:** Library representatives and members of the public, who are invited to answer designated surveys.
 - Require Swedish interface.
 - Estimated max. 20.000 active survey respondents during first year.
 - Assuming each of 5.000 libraries will need at least one staff person to respond to national surveys.
 - Additionally, approximately 200 libraries are assumed to conduct local surveys with an average of 75 respondents each. I.e. 200 times 75 equals 15.000 respondents.
- **Statistical Analysts:** Trained library representatives and other professionals who select and analyze statistical data.
 - The analysis is performed by means of the central analysis web application or data is exported to local tools, e.g. Excel or compatible spreadsheets programs.
 - Require Swedish interface.
 - Estimated max. 10.000 active statistical analysts during first year.
 - Assuming each of 5.000 libraries will require at least one staff person to examine local and national statistical data.
 - Additionally, 5.000 administrators, researchers and general members of the public are assumed to look at statistical data from the system.
 - The majority of users are assumed to check a few statistical reports and indicators – i.e. only a minority of users will execute advanced multi-dimensional analysis or extensive data mining functions.

3.3. User Functions

A list of nineteen anticipated user functions was drawn up based on preliminary analysis and initial contacts with a few potential users.

The proposed functions were subsequently explained to user representative, who were asked comment and to assign a relative priority to each function. The priorities were designated *High, Medium or Low*.

The proposed functions together with the users' priorities are tabulated below.

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
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

The user's priorities were used to compute a weighed score for each function.

The scores were calculated by this formula: $3*High + 2*Medium + Low$.

Function *f19* is for example assigned the following score:

$$3*16 (H) + 2*5 (M) + 3 (L) = 61$$

The functions and priorities are illustrated below sorted by descending score values.

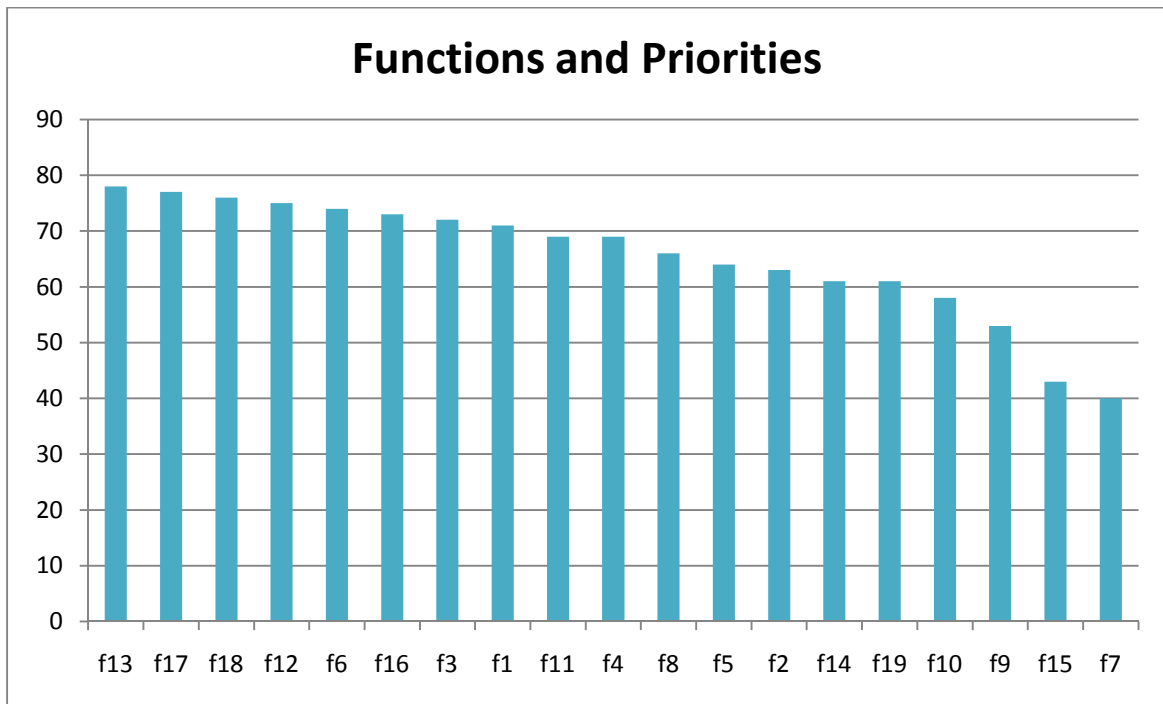


Fig. 3 Functions and priority scores

The user representatives did not put forward any major additional functions and apparently considered most of the proposed functions to be pertinent.

The average priorities across all functions and respondents are 63% *High*, 20% *Medium*, 5% *Low* and 12% *Unknown* (?).

The most notable exceptions are functions number fifteen and seven: Only 14 percent of respondents consider function *f15 Restricted access to data* a high priority and 46% did not assign any priority to function *f7 Xml documents and schemas*.

The limited interest in restricted access to data is a potential boon: It is tricky to prevent determined perpetrators from extracting classified information from statistical databases by means of clever cross-queries etc. Avoiding data access limitations could significantly simplify the system design.

4. GENERAL SYSTEM DESCRIPTION

4.1. System Context

The new system, which is tentatively called *BibStat*, includes a set of software tools and services to collect, manage, analyze and disseminate statistical information and performance indicators related to Swedish libraries. Measurements data must be described and identified by corresponding metadata.

BibStat complements existing tools to collect or analyze library performance data and also provide facilities to analyze transaction data imported from designated local library systems or Electronic Content Providers.

BibStat can import data uploaded from local systems and interacts directly with third party application via web services with industry standard system interfaces.

The core of the system is a logical database service or *Repository* to store and retrieve statistical measurements, transaction data, survey responses and other analytical data. The Repository is based on a relational database system with general data management features.

The primary external system interfaces include a set of on-line *Web Services* that are accessed by distributed client applications via standard Internet protocols.

The BibStat system also includes a number of *Web Applications* with interactive human user interfaces that are accessed with standard browsers via the Internet.

The system can handle shared and local statistical measurements, surveys and transaction data. An interactive web application is used to maintain metadata, which describe the different types of measurements data etc.

The system can **input** data via four different logical channels:

- **Statistical measurements** and parameters are input via an interactive web application
- **Survey responses** are collected via an interactive web application
- **Transaction data** extracted from local library applications are imported via a central web service.
 - Manual upload of transaction datasets is handled via an interactive web application
 - Periodic automatic upload of transactions is performed via a central web service
 - The central *Import Data* web service utilizes industry standard internet protocols, e.g. *AtomPub*²²
- **Usage Statistics** are harvested directly from local Content Provider web services

²² IETF RFC 5023 The Atom Publishing Protocol. <http://www.ietf.org/rfc/rfc5023.txt>

- The central Harvest Data Client utilizes industry standard internet protocols, e.g. *SUSHI* ²³

The system can **output** data two ways:

- An interactive **web application** is used for on-line statistical analysis and reporting
- A system **web service** exports data to third party tools and applications, e.g. *Excel* or local reporting systems
 - The central *Export Data* web service supports industry standard internet protocols, e.g. *XMLA* ²⁴ and *AtomPub* ²²

The following system overview identifies the major features and information flows. This figure does not necessarily reflect the internal system architecture or distribution of functions.

²³ ANSI/NISO Z39.93 - The Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol. http://www.niso.org/kst/reports/standards/kfile_download?id%3Astring%3Aiso-8859-1=Z39-93-2007.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hPft48o3jFRwNR8LAUv7bDUIEAytQp1kLJGhhtjwcl3V

²⁴ XML for analysis. <http://www.xmla.org/index.htm>

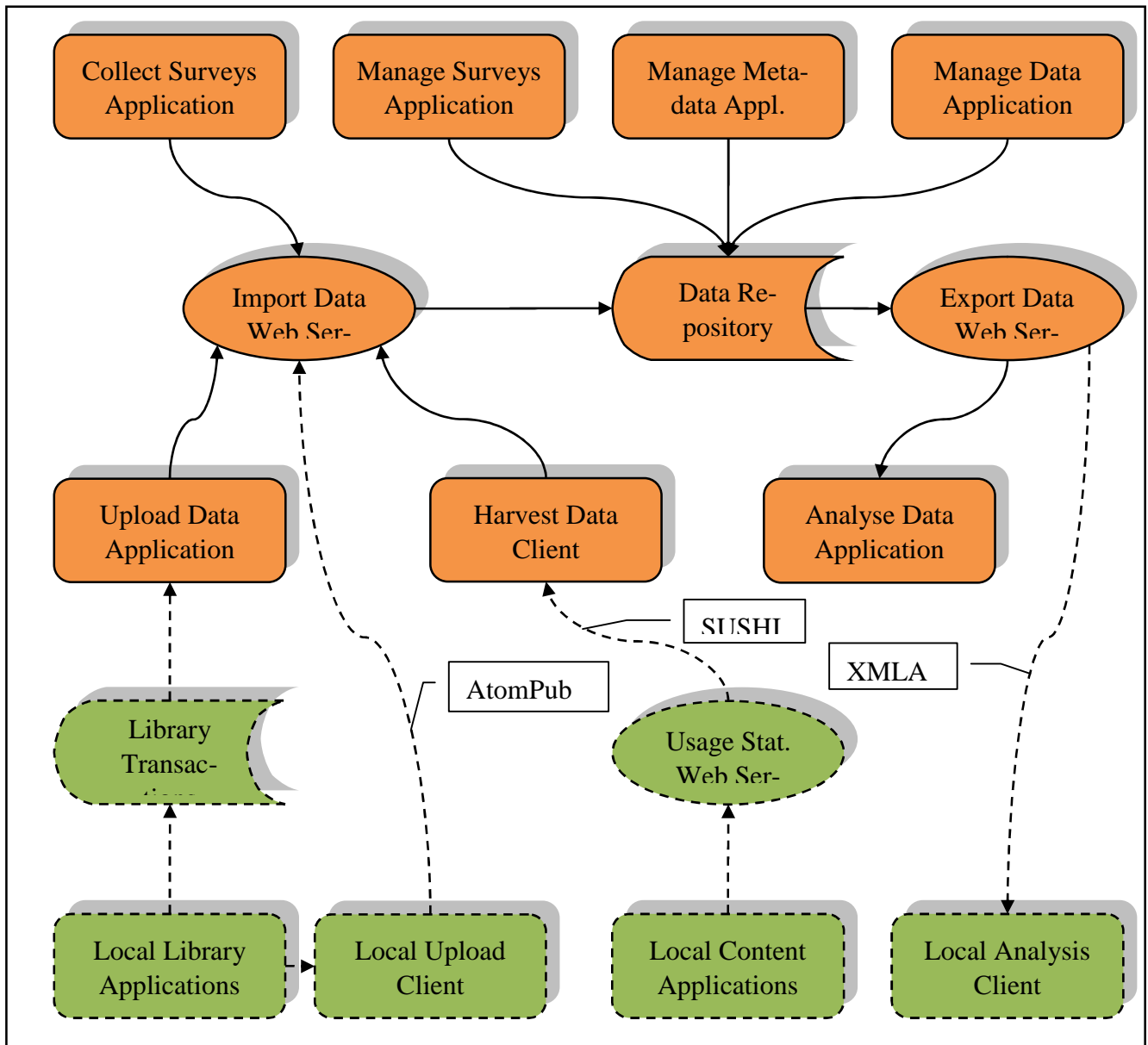


Fig. 4 System Overview

Solid lines designate components that are part of the central system, while external facilities and component are identified by dotted lines.

4.2. Major System Features

The system contains the following major features:, which are specified by the functional requirements.

- **Data repository.** Data and metadata is stored within a central logical database or repository. The repository provides general relational database facilities augmented by special functions to support multi-dimensional data analysis, ROLAP²⁵
- **Metadata Management.** The system can handle different types of statistical measurements, transactions and usage data. An interactive web application is used to manage metadata which facilitate data retrieval and analysis. Metadata describe data structures, domains, attributes and logical relationships as well as information about the authorship, origins and methodology used to collect the data.
- **Data Management.** Different types of statistical measurements and attribute values are entered and updated via an interactive web application.
- **Survey Management.** Statistical data is collected via electronic surveys, where participants are invited to submit designated information via customized interactive web applications.
- **Survey Collection.** Responses to surveys are input via an interactive web application with user authentication, data validation and branching logic.
- **Transaction Upload.** Transaction data from local library systems are uploaded via an interactive web application. Customized local applications may alternative submit transactions directly via a central web service. The *Import data* web service utilize standard Internet protocols (*AtomPub*²²)
- **Usage Data Harvesting.** Usage data are harvested from Electronic Content Providers via standard local web services (*SUSHI*²³). Data from standardized usage reports are imported to the central Data Repository database, where the data can be extracted for further analysis.
- **Data Analysis.** A central interactive web application provides functions to select and analyze data-including cross tabulation and standard reports.
- **Data Download.** Analytical data is downloaded from the central Repository database to local analysis tools from the central Repository database. Data can also be extracted directly to Excel and other local analytical tools via a central web service. The *Export data* web service supports standard internet protocol (*XMLA*²⁴).

4.3. Major System Constraints

The system must support the reporting needs for all of the publicly funded libraries in Sweden, i.e. a total of more than 7.000 institutions. The libraries range from small operations staffed by a single person to large organizations with numerous branches.

²⁵ Relational Online Analytical Processing.

http://searchoracle.techtarget.com/sDefinition/0,,sid41_gci214582,00.html

The specified functions must be free to use by designated library representatives, i.e. without usage charges or license fees to third parties. This does not preclude any charges or fees levied by the public authorities, which assume responsibility for the system.

The authority that assumes future responsibility for Swedish national library statistics may wish to introduce some additional conditions or dependencies, which affects the system design, implementation strategy or operational setup.

4.4. Operational Scenarios

The present specifications do not stipulate any specific implementation strategy or operational scenario.

A combination of third-party hosted services integrated with modules operated by the future designated authority could for example represent a viable solution.

Smaller libraries may prefer to rely entirely on external services, while larger library organizations have the necessary IT resources to integrate local applications with the central web services.

Regardless how and where different functions are hosted physically, it is important, that libraries are provided with the means to control and manage their own private data, metadata and surveys within the framework of the shared system.

5. FUNCTIONAL REQUIREMENTS

The following functional requirements are designed to support the designated user requirements and system features.

System designers may propose to satisfy the user requirements by means of different but equivalent functions.

5.1. Manage Surveys

The Survey Management facility includes functions to specify survey questions, adapt a survey response application, validate survey responses and save the survey results.

5.1.1. Specify Survey Content

Define-Survey-Content is a function for survey managers. The purpose is to specify the questions, validation rules and other information about a survey.

1. A survey author must specify the questions and other characteristics of a survey.
2. Each survey must include a list of questions to be answered by survey respondents.
3. A survey question can request a data value, i.e. *measurement* of something. E.g. “How many patrons visited your library in 2009?”
4. A survey question can request a metadata value, i.e. *dimension* identifying some measurement. “E.g. what is the location of your library?”
5. Each answer can be restricted to a predefined range of values, i.e. a numerical interval or a list of options. E.g. what type of library do you represent? Please select *Public* library, *Research* library, *Hospital* library or *School* library
6. Each question can be specified as mandatory, i.e. default is optional.
7. Surveys authors may specify conditional branching logic based on answers to preceding questions within a survey.
8. A survey author can update survey questions etc. for surveys without any survey responses.
9. A survey author can delete an entire survey including any responses to the survey.

5.1.2. Customize Survey Format

Customize-Survey-Format is a function for survey managers. The purpose is to customize the Input-Survey-Response function to support a given survey.

1. The survey author may specify some introductory information, which is displayed at the beginning of the designated survey.
2. Survey author may specify a password prompt or other suitable authorization method, so that only authorized users can answer the designated survey.
3. Starting and closing times may be specified for a survey. I.e. users can not answer the survey before the starting time or after the closing time.

4. The maximum number of responses may be specified for a survey. I.e. the survey is closed, when the designated number of responses has been collected.

5.1.3. Manage Survey Invitations

The *Manage-Survey-Invitations* function is used by survey managers. The purpose is to select designated persons to participate in a survey, distribute invitations, track responses and distribute reminders.

1. Each survey author can maintain lists of users, which can be invited to participate in surveys.
2. Groups of persons can be selected to participate in several surveys.
3. Survey authors cannot access users and invitations belonging to other survey authors.
4. Data about persons can be imported to a survey authors list of invitees.
5. A survey author can interactively create, update or delete information about survey participants.
6. Groups of participants can be selected and invited via customized e-mail messages or documents.
7. Individual authorization credentials can be distributed together with survey invitations.
8. Survey responses from identified persons can be linked to invitations.
9. Receipts and thank-you notes can be distributed to known survey respondents.
10. Reminders can be distributed to invited participants who have not responded.

5.1.4. Manage Survey Responses

1. An authorized user can review, correct or delete previous survey responses that are waiting for further processing.

Manage-Survey-Responses are an interactive web application for survey authors. The purpose is to check for duplicates or obviously incorrect responses etc.

1. A survey author cannot access responses to surveys created by other survey authors.
2. A survey author can update answers within a survey response.
3. A survey author can delete individual or all responses to a survey.
4. Raw survey responses can be mapped to shared statistical measurements and attributes.
5. Survey responses can be mapped to formats and structures used by the Survey Analysis functions.

5.2. Collect Surveys

Collect-Survey-Responses is a function used by survey respondents. The purpose is to enter the answers to a given survey. An authorized user can also correct or delete a response, while it is waiting further processing.

Customized web forms to collect responses to a given survey are generated on basis of the designated survey specifications.

6. Users can access a given survey from a web browser via a URL linking directly to the designated survey.
7. Users must be authenticated in order to access a survey, which requires authorization.
8. The system must prevent duplicate survey responses from an authenticated user.
9. The user interface must be intuitive, so that normal computer-literate users can submit and update a survey response without prior formal training.
10. System must provide relevant and useful help facilities on demand.

5.3. Manage Metadata

Manage -Metadata is a function for metadata managers. The central database repository includes several types of metadata information.

1. **Syntax and format** of datasets and measurements including validation rules and functional dependencies. This information is specified by means of xml schemas and logical database schemas.
2. **Origin and provenience** of datasets and measurements. This information is specified by means of general shared hared attributes, e.g. origin, time, location etc. supplemented by proprietary descriptions of methodology and other pertinent information.
3. **Relational model** that describe the multi-dimensional relational structures and tables, functional relationships and keys
4. **Data Dictionary** that describes the context, meaning, attributes hierarchies and synonyms of data elements, statistical measurements and dimensions. The data dictionary may internally utilize the industry standard *OWL Web Ontology Language*²⁶

The purpose is to specify data elements, semantic relationships between measurements and dimension attributes that describe the measurements.

Metadata relations are used for different types of statistical analysis.

1. Survey *measurements* are specified with reference to the objects or concepts being measured. E.g. *Number of user workplaces*. Several surveys may reference the same measurements.

²⁶ W3C OWL 2 Web Ontology Language. <http://www.w3.org/TR/owl2-overview/>

2. Measurement attributes (*dimensions* identify the measurements. E.g. *Library Name* or *Measurement Period* identifies, where and when the aforementioned *Number of workplaces* were measured or counted. Several surveys may reference the same dimension attributes.
3. Dimension attribute values can be specified as *hierarchies*, which are used to aggregate measurements. E.g. Geographical and organizational entities such as *City < Region < Country* or time periods such as *Day < Week < Month*. Hierarchical relationships are implicitly shared qua the participating dimension attributes.

5.4. Analyze Data

Analyze -Data is a function for statistical analysts. The purpose is to perform basic on-line statistical analysis including selection, cross-tabulations and calculation of standard statistical indicators.

1. List surveys ordered by relevant parameters such as survey names, survey authors, survey dates or number of responses.
2. Search for survey questions by relevant parameters such as question texts, measurement types or dimension types.
3. Display Frequency Analysis of answers to a selected survey question. E.g. *Number of employees*.
4. Display standard statistical parameters such as Mean, Median, 95% Confidence Interval, Standard Deviation and Standard Error of answers to a selected survey question. E.g. Mean of *Operating expenditure* divided by *Number of loans*.
5. Display Cross Tabulation of answers that are related, i.e. share common dimension attributes. E.g. *Number of electronic collections* versus *Region*.
6. Specify *OLAP Cubes*²⁷ with reference to metadata.
7. Display *OLAP Slices* and *Dices*.
8. Display and *rotate pivot tables*.
9. Export analytical data to Excel or other compatible third party applications.

5.5. Upload Transaction Data

The Upload Transaction Data facility includes functions to import and datasets of transactions extracted from local IT systems, e.g. loan transactions or holdings records.

5.5.1. Specify Transaction Datasets

Specify-Transaction-Datasets is a function for upload managers. The purpose is to specify the format and semantics of transaction records.

1. The syntax (format) of transaction datasets must be specified with reference to XML schemas²⁸ supported by Excel²⁹ and compatible tools.

²⁷ On-Line Analytical Processing terms. <http://www.olapcouncil.org/research/resrchly.htm>

²⁸ W3C XML Schema Part 1. <http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/structures.html>

2. The semantics (meaning) of transaction data elements must be specified with reference to the same types of metadata as used for other data.
3. Different transaction data elements and survey data can share metadata specifications.

5.5.2. Upload Transaction Datasets

Upload-Datasets is a function used by upload managers. The purpose is to transfer files with transaction data from a local PC to the central system.

1. Select local data file and upload to central system.
2. Uploaded data files must conform to formats exported by Excel ³⁰
3. Upload managers can delete and replace uploaded datasets.
4. Datasets can be designated as restricted, i.e. default status is public.

5.5.3. Upload Data Clients

Local web service client applications can periodically submit data directly via the central *Import Data* web service.

Security considerations may dictate restrictions to the use of this function by local clients.

5.6. Harvest Usage Data

Harvest-usage-data is a function used by upload managers. The purpose is to fetch and save *COUNTER usage* data ¹⁶ harvested via SUSHI ²³ compatible web services provided by distributed content providers.

This function may be set up to harvest designated data periodically, i.e. without manual intervention.

1. Specify Content Provider web service
2. Specify usage data
3. Get usage data from content provider web service

5.7. Communication Interface Requirements

The system must comply with the following data communication standards:

- Web applications must support general Internet standards adopted by widespread browser applications including http/https and *XHTML*.
- Local web service clients that export transactions to the central **Import Data** may utilize the *Atom Publishing Protocol* ²² (AtomPub)
- Local web service clients, that import data from the central **Export Data Service** should support the *XML for Analysis web service standard* ²⁴ (XMLA) including the *Multi-dimensional Expressions* standard ³¹ (MDX)

²⁹ Excel XSD support. <http://office.microsoft.com/en-us/excel/HP102064141033.aspx?pid=CH100648521033>

³⁰ XML in Excel. <http://office.microsoft.com/en-us/excel/HA102063961033.aspx?pid=CH100648521033>

- Local Content Provider applications should support the *Standardized Usage Statistics Harvesting Initiative*²³ (SUSHI) web service standard and the *Counting Online Usage of NeTworked Electronic Resources* (COUNTER)¹⁶ codes of practices.

³¹ MDX / mdXML. <http://www.xmlforanalysis.com/mdx.htm>

6. NONFUNCTIONAL REQUIREMENTS

Nonfunctional requirements depend on the sort of system architecture, implementation strategy and operational scenario selected by the authority, which will be responsible for the system. Specification of nonfunctional requirements, before the new authority has been appointed, is therefore problematic.

The following aspects should be addressed by the detailed systems design and could also be covered by Service Level Agreements:

- Usability
- Performance
- Reliability
- Security
- Supportability
- Infrastructure
- Implementation

7. SUMMARY AND RECOMMENDATIONS

The requirements survey among library representatives confirmed a broad consensus about the proposed assumptions and priorities for the user functions.

The system should be designed around a relational database repository with management information, statistical measurements and transaction data coupled with an integrated data dictionary.

Interactive web applications should provide facilities to manage surveys and metadata as well as to maintain and analyze statistical information.

System web services with standard interfaces should facilitate integration with local library applications and with third party services to conduct public surveys, to harvest usage data and to analyze aggregated information.

7.1. User Requirements

The most significant finding is that libraries desire to aggregate and analyze private types of statistical data – in addition to common national surveys.

Libraries also expect to collect different sets of transaction data exported automatically from library applications – in addition to plain statistical measurements.

Another noteworthy result is that libraries favor submitting transaction data without human intervention direct from local client systems – instead exporting and uploading files manually.

The majority of respondents prefer to download data to local spreadsheets etc.; but also call for shared analysis and reporting functions.

7.2. System Overview

The new system includes interactive web applications so that users can manage surveys, collect statistical information, specify metadata and analyze statistical data.

The system also provides web services that can communicate directly with third party applications. Central web services should support standard internet client-server protocols e.g. for data selection and navigation, harvesting of usage data and data analysis.

The statistical data and transactions must be stored, organized and maintained within a general database repository, so that the information can easily be searched, analyzed and exported - practically forever.

7.3. Metadata dictionary

The requirement to handle different proprietary types of statistical data elements represents a significant technical challenge: While it is unproblematic to document and

process a limited number of shared data types; it is more difficult to handle a rising number of private data types.

In order support retrieval and analysis with private statistical data types, it is of primary importance to design a flexible and user friendly data dictionary with comprehensive metadata.

The metadata must include relevant information about the source and origin of statistical data as well as their structure and syntax, calculations, validation rules, logical relationships and semantics.

7.4. Repository Database

The database repository utilizes data management and utility functions provided by a general relational database management system.

The shared database includes information about surveys, users, and systems management as well as data dictionary information, statistical data and transaction datasets and analytical structures.

7.5. Survey Management

Several vendors offer flexible and inexpensive services to conduct public surveys via interactive web applications. Collecting survey responses via e-mail forms may look like a tempting alternative, but often incur technical and practical problems.

Many popular services offer limited facilities to invite designated respondents via e-mail, but the new system needs more advanced integrated functions to manage survey invitations and responses from thousands of users.

7.6. Data Management

Statistical data and metadata is maintained via interactive web applications integrated with the data dictionary facilities.

7.7. Data Import

Data from library applications can be uploaded to central web services from customized local systems, i.e. "*pushed*" from local client applications.

Alternatively, usage data and other transaction data can be harvested by standard central web service clients, i.e. "*pulled*" from local servers.

7.8. Data Analysis and Export

Users may select and analyze Statistical information via a web application with standard reporting functions.

Spreadsheets and other tools can also access data directly via standard system interfaces to the database repository.

7.9. Next Steps

The metadata design and database structures constitute the foundation of the system while the web services represent the primary infrastructure. It would be difficult to alter the fundamental data structures and system interfaces later, whereas a sound data model together with well designed system interfaces can facilitate future additions of third party applications and services.

It is also helpful to design, test and modify the logical data models, metadata structures and external web service interfaces before detailed system design and implementation. Consequently, the first task should be to design the metadata and logical structures so that both shared and private types of statistical information can easily be stored, maintained, searched, retrieved and analyzed.

The next step could be to design the external web services and to test their system interfaces by limited prototypes together with a few relevant local library applications.

The metadata and data design could also be tested with functional prototypes to store and retrieve different types of statistical measurements and attributes – preferably collected by selected libraries.