Chapter 4

CDS-Invenio Digital Library Software Evaluation
http://cdsware.cern.ch/invenio/index.html

Introduction

CDS-Invenio is a web based integrated digital library system developed by CERN. CERN Document Server Software (CDS-Invenio) is a suite of applications which provides the framework and tools for building and managing an autonomous digital library server. CDS-Invenio uses entirely freeware technology and it is available under the terms of GNU General Public License. It is developed by the CERN Document Server team and is validated by the CERN Scientific Information Service.

CERN, the international European Organization for Nuclear Research has been involved since its early beginnings with the open dissemination of scientific results. It is world’s largest particle physics centre. CERN Document server has now two software which are made available under open source license. CDS Invenio and CDS Indico. CDS Invenio is digital library software and CDS Indico is the conference management software.

CDS Invenio helps to run electronic preprint server, an online digital library catalogue or a document repository on the web. It complies with the Open Archives Initiative (OAI) and uses MARC21 as its underlying bibliographic standard.

It is a powerful, flexible digital library system suitable to handle very large collections of full text documents. It tries to combine the best of the traditional library world and modern
information retrieval technology. It is a user based digital library software. All the jobs in the repository are carried out by different users with different roles. CDS-Invenio has grown into a large software suite intended to cope with large collections (almost 1 million records at CERN), and with advanced library-type functionalities.

CDS-Invenio package can be used either as a general document management solution, a library system or an institutional repository in parallel with the production of a digital library system.

The design of CDS-Invenio is different as well as different terminologies are used in CDS-Invenio hence it takes time to understand the functionality of the software. The software is specially designed to cope with very large repositories. All modules provide web-based administrative tool kit. The key feature of CDS-Invenio’s architecture is its modular logic[1]. Each module carry out a specific, defined, functionality of the digital library system.

The software has the following modules which perform different functions which is one of the unique feature of CDS-Invenio.

**WebSubmit:** Submitting data

**BibHarvest:** Harvesting OAI repository

**BibConvert:** Harvesting non-OAI collections

**BibFormat:** Formatting and linking records

**WebSearch:** Searching metadata/citations/full text

**BibWord:** Indexing metadata and full text

**WebAccess:** Managing complex collection hierarchy

**WebPerso:** Personalizing web access

**BibData:** Modifying records (librarians only)

**BibSched:** Takes care of the task scheduling and the task execution
First version of CDS-Invenio was released during August 2002 with version number as 0.0.9. For the present study installation of 0.92.1 version was carried out and selected for evaluation study. This version was released on 20th February 2007. The latest version of CDS-Invenio is 0.99.0 which is released on 23rd March 2008. The following sections lists different observations after having installation of version 0.92.1.

## 4.1 Content Acquisition

### 4.1.1 Which document types can be added in the software (books, reports, journal articles, lecture notes, technical reports, thesis, images, audio/video files, data set files etc.?)

CDS-Invenio supports to allow to add bibliographic data of books, reports, journals as well as all other types of digital media such as text, audio, video, image, data set files etc.

### 4.1.2 Does the software have capability to define the domain/scope of Digital Library for whom it is intended for?

Yes, CDS-Invenio allow to define the domain/scope of the digital library for whom it is intended for.

### 4.1.3 If yes, what level of users it is intended for?

CDS-Invenio can be used to organize any type of digital documents. It is designed for all types of libraries.

### 4.1.4 Which digital document file formats does the system support (proprietary as well as open source)?

CDS-Invenio has the ability to accept documents in all desired formats. Documents can consist of multiple files. CDS-Invenio gives the ability to submit a document with more than one file. Thus, it is possible to submit text with a separate file for accompanying tables. The system administrator can limit the formats of submitted documents. This allows the repository to define a policy according to which it accepts specific formats of digital objects that it can manage from a technical point of view.
4.1.5 Does the system carry out duplicate checking while uploading the data? What happens if the same digital document is submitted multiple times?

No, CDS-Invenio does not carry out duplicate checking. If the same digital document is submitted multiple times it gets added into the repository.

4.1.6 How does the system allow to upload digital objects, whether normal digital objects or compressed digital objects?

CDS-Invenio supports to upload digital objects in normal digital formats as well as compressed digital formats (such as zip, tgz, gz, gzip etc.).

4.1.7 Does the software support to upload digital objects from existing URL? No

4.1.8 Does the system support distributed/community based acquisition?

Yes, if end user has login name and password he/she can submit document through web interface.

4.1.9 Does the software support metadata/content import? In which file format data is required for importing?

Yes, CDS-Invenio support metadata/content import. The data has to be in MARC XML to MARC21 or Dublin Core (DC) format. BibHarvest and BibConvert tools allow to run massive importation of records from either OAI compliant data provider or from non OAI complaint provider. Full texts files have to be separately imported.

4.1.10 Does the system support importing of bulk metadata/content of digital objects?

Yes, CDS-Invenio support importing of bulk metadata of digital objects. Full texts files have to be separately imported.

4.1.11 Does the software support import of automatic metadata for the digital objects that are added into the repository?

The system supports automatic extraction of "keywords" from the full text of the documents based on the frequency of the specific terms.

4.1.12 Does the software support metadata/contents export? In which file format data is exported?
Yes, CDS-Invenio supports metadata export in BibTeX, Dublin Core, EndNote, HTML brief, HTML detailed, MARC, MARCXML, NLM, photocaptions only, portfolio format. The full texts are not exported at present.

4.1.13 **Does the system support export of bulk metadata/content of digital objects?**

Yes, CDS-Invenio support to export bulk metadata of digital objects.

4.1.14 **Does the system allow adding past versions of the digital document files?**

Yes, CDS-Invenio supports adding past versions of each digital document to control different editions of registered objects.

4.1.15 **Does the software support its own accession number for each document that gets added into the repository?**

Yes, CDS-Invenio support its own accession number for each document that gets added into the repository.

4.1.16 **It is necessary to know whether software support weed out policy?**

Yes, CDS-Invenio support weed out policy. Documents that have been erased by CDS-Invenio are moved to "non-defined collection" that means if a collection or an item does not appear during the search it does not mean that it has completely disappeared from the repository but has been kept in the archive directory of CDS-Invenio for future actions.

4.1.17 **Does the system support to delete items from the collection/repository, move objects from one collection to other collection to other collection, cross-list objects across different collections?**

Yes, CDS-Invenio support to delete items from collection/repository as well as move objects from one collection to other collection, i.e content mapping.

4.1.18 **Which persistent identifier scheme does the system support?**

CDS-Invenio does not support any global persistent identification scheme. It has its own persistent ID number for all documents that are added into the repository.

4.1.19 **Does the system’s persistent URL have capability to handle an object’s change in location and state in future with same identifier number?** No
4.2 Content Management

4.2.1 Does the system have proper work flow in place which handles different submission processes?

Yes, CDS-Invenio has very good work flow which handles different submission processes.

4.2.2 If the system supports to have proper work flow for all actions into the repository then whether submission supports following options:

4.2.2.1 System allows to create overall system administrator Yes
4.2.2.2 System allows to submit documents by users via web Yes
4.2.2.3 System allows to set up submission rules Yes
4.2.2.4 System can define accept/reject policy for all documents that gets added into the repository Yes
4.2.2.5 Allows multiple collections within same installation of the system Yes
4.2.2.6 Home page for each collection/community Yes
4.2.2.7 Allows to set different policies for different collections Yes
4.2.2.8 Allows to set different authorization policies for different collection/items added into the repository Yes
4.2.2.9 Allow to edit metadata submitted by users Yes
4.2.2.10 Sends email notification to users/submitters/authors Yes
4.2.2.11 Sends email notification to metadata reviewers Yes
4.2.2.12 Sends email notification to reviewer Yes
4.2.2.13 Sends email notification to administrators Yes
4.2.2.14 Allow users to review completed content Yes
4.2.2.15 Allow users to review uncompleted content Yes
4.2.2.16 Allow content administrator to review submissions Yes
4.2.2.17 View pending content administration tasks Yes
4.2.3 Does the system support to upload documents into the repository through Librarian’s interface as well?

No, CDS-Invenio does not have a separate librarian’s interface all actions in CDS-Invenio are carried out through Web Interface.

4.2.4 Does the software provide easy way for adding/editing of records for Librarian as well as for the remote user? Are insertion and deletion of records easy? Yes

4.2.5 Are the supported file formats well documented in the system? Can new file formats be added or removed by the end user?

The software does not have any mention about the restriction of file formats. It can upload any type of file format.

4.2.6 Does the repository software have any capabilities, such as access control lists, Internet address filters, etc., that limit who is allowed to submit items in the software?

All the operations in CDS-Invenio are carried out through login and password hence only those members can use the repository who have access to login ID and password.

4.2.7 Does the repository software maintain audit logs that identify by whom and when documents are submitted into the repository? No

4.2.8 Does system support knowledge organization systems such as ontologies, thesauri, classification systems and taxonomies for users while adding documents via web in the repository?

Yes, CDS-Invenio supports to organize all documents into the repository using thesaurus.

4.2.9 Does the system support digital objects to be grouped into organized structure (i.e. linear/hierarchical/graphical structure)? Yes

4.2.10 Does the system support to generate authority files? No

4.2.11 Does the system allow digital object to be a member of multiple collections i.e. content mapping?
Yes, In CDS-Invenio a single record can appear in several collections. There is no limitation to the number of collections in which a record can appear. Virtual mapping is possible in CDS-Invenio.

4.2.12 Does system allow users to modify (add/delete/update) metadata/complete document added within the repository? Yes

4.2.13 Does the system support to show strength of each collection? Yes

4.3 Metadata Submission and Support

CDS-Invenio’s basic metadata scheme is MARC21 as well as it supports creation of basic Dublin Core (DC)[2]

Since basic Dublin Core, is sometimes inadequate to satisfy the needs of a specific organization, OAI supports the parallel use of metadata sets, allowing repositories to use additional metadata specifically for each specific collection.

CDS-Invenio allows the repository administrator to predefine fixed values in metadata fields to simplify their inclusion.

A document type can be one of "Thesis", "Photos", "Videotapes"... or any other type. A document type is always defined by its metadata. It may or may not have a full text file attached to it i.e. end users can create bibliographic database as well using CDS-Invenio.

Metadata in cds-invenio is performed by automated and semi-automated procedures of document harvesting (BibHarvest module) - by applying either standardized approaches such as using OAI-PMH compliant metadata harvesting or ad-hoc procedures such as shallow web harvesting[3].

4.3.1 Does the software support to add descriptive metadata (such as author, title, subject, publisher, copyright, year of publication etc.) for each digital object that gets added into the repository? Yes

4.3.2 Does the software support to add administrative metadata (such as technical aspects of digital documents, source of information(when and how it was created), rights manage-
ment, what is file size, in what file format digital document is in, what is needed to view the digital document) for each digital object that gets added into the repository?

Default templates are not available but one can add administrative metadata in CDS-Invenio.

4.3.3 Does the software support to add structural metadata (such as description, owner, data type, date deposited, version number, date of last revision, how compound objects are put together, how pages are ordered to form chapters) for each digital object that gets added into the repository?

Yes, CDS-Invenio supports to add structural metadata.

4.3.4 Does the software support adding/editing/deleting metadata fields? Yes

4.3.5 Are help messages given in each metadata field while entering data? Yes

4.3.6 Does the metadata entry field has one field mandatory? What are the mandatory fields in each software?

Yes, CDS-Invenio has 'Title' and 'Author' as a mandatory field.

4.3.7 Does the software support Unicode character set for metadata entry?

Yes, CDS-Invenio supports UTF-8.

4.3.8 Does the software allow to create and manage variety of different metadata schemas to meet different needs across the subject disciplines?

CDS-Invenio’s basic metadata scheme is MARC21. It also supports the creation of basic Dublin Core (DC). The Open Archive Initiative (OAI) protocol additionally offers the ability to search even in non-certified DC fields (that is, additional fields not part of basic DC).

4.3.9 Can the metadata fields or formats be customized? Yes

4.3.10 What type of Metadata Schemas are supported by the software as a default?

- DC (Dublin Core) Yes
- MARC 21 Yes
4.3.11 Does the software have default metadata entry templates? Yes

4.3.12 Are the metadata formats documented?
CDS-Invenio have documentation of MARC 21.

4.3.13 Does the software support different interface for metadata entry and whether customized data gets added into the system? Yes

4.3.14 How does the software verify that submitted objects and metadata are correct?
There are no ways by which it can verify that submitted objects and metadata are correct.

4.3.15 Are there automated checks of the metadata, such as verifying that a date entered into a field is really a date string? No

4.3.16 Does the repository software have any means to verify that objects or metadata have not been tampered with, such as checksums or digitally signed checksums? No

4.3.17 Does the software support real time updating and indexing of accepted contents?
Yes, CDS-Invenio support real time updating and indexing of accepted contents.

4.3.18 Does the software allow for metadata extensibility and complexity? Yes

4.3.19 Does the software support metadata versioning? What happens to old versions of metadata or content objects when they are replaced by new versions or modified in some way? No

4.3.20 Does the repository software support provenance metadata? What is included in the provenance?
Yes, CDS-Invenio supports MARC 21 hence all metadata details can be added into the system. Provenance can include details about the origin of the document etc.

4.3.21 Does the software support metadata crosswalk? No

4.3.22 Does the software support "thesaurus" building? Does it specify which thesaurus standard is used? Which subject heading DDC, DC or LC is used for thesaurus building?
Yes, CDS-Invenio helps in thesaurus building. CDS-Invenio can use any subject heading for building thesaurus.
4.3.23 Does software have ontology support? No

4.3.24 How does the repository verify file types, for example, if JPEG images are submitted to the repository, does the software verify that it really is a valid JPEG, or does it just blindly take the submitter's word? No

4.3.25 While exporting metadata whether software support checksums? No

4.3.26 It is necessary to know whether the software support RAP protocol? Yes

4.4 Classification

4.4.1 Does the system allow digital objects to be grouped into classification/subjects? CDS-Invenio supports entering of class numbers.

4.4.2 Which classification system does the software support? It supports to enter any class number such as Library of Congress/UDC/DDC/BC etc.

4.4.3 Does it allow to group documents as per the class number? Yes

4.4.4 It is also necessary to know whether system support hierarchical browsing of collections through class number? Yes

4.5 Information Search & Retrieval

4.5.1 Browsing

4.5.1.1 Author/Creator/Contributor Yes

4.5.1.2 Title of the document/Title of the book/Title of the article etc. Yes

4.5.1.3 Issue Date/Date of Publication Yes

4.5.1.4 Collection Yes

4.5.1.5 Communities No
4.5.1.6 **Subject browsing**  Yes

4.5.1.7 **Publisher wise browsing**  Yes

4.5.1.8 **Table of Contents browsing**  No

   The system allow users to search and browse against any metadata field that is defined by the user in the system as it uses MARC 21 format.

4.5.1.9 **Multi-Dimensional Browsing**  No

### 4.5.2 Searching

4.5.2.1 **Searching content via different metadata fields such as author, title, subject, publisher etc.**

   Yes, CDS-Invenio supports to search documents added into the repository by any type of MARC 21 data field.

4.5.2.2 **Full text searching**  Yes

4.5.2.3 **Boolean (AND, OR, NOT) searching**  Yes

4.5.2.4 **Basic search**  Yes

4.5.2.5 **Advanced search and retrieval feature**  Yes

4.5.2.6 **Does the system allow to search the database putting boolean operators such as !, |, & etc. as well as ”must contain”, ”should contain”, ”may contain”, ”not contain” as operators in place of AND, OR, and AND NOT Boolean operators respectively.**

   CDS-Invenio supports all boolean operators such as AND, OR, NOT as well as +, -, ! as well as must contain, should contain, may contain.

4.5.2.7 **Truncation/Wild card searching:**

   Yes, CDS-Invenio supports searching via truncation searching by using operator such as *.

4.5.2.8 **Exact words/phrases searching:**  A query may be entered in quotes to search for an exact match of the phrase.

   Yes, CDS-Invenio supports to search by exact phrase as well as by partial phrase.
4.5.2.9 Proximity searching Yes

4.5.2.10 Stemming Yes

4.5.2.11 Fuzzy searching Yes

4.5.2.12 Phonetic searching Yes

4.5.2.13 Case sensitive or case insensitive while searching

   Yes, CDSInvenio supports searching based on case sensitive.

4.5.2.14 System should support term weighting Yes

4.5.2.15 System should support search history option, reuse of query, query save option Yes

4.5.2.16 Boosting the term Yes

4.5.2.17 Range searching Yes

4.5.2.18 Use of thesaurus or permuted index for searching No

4.5.2.19 Expand search Yes

4.5.2.20 Lateral searching No

4.5.2.21 Search support for special collections No

4.5.2.22 Searching via table of content and classification codes No

4.5.2.23 Ability to browse subject/author authority files No

4.5.2.24 PIC variant searching (plurals, international spelling variants, and compound words) No

4.5.2.25 Support multilingual search and retrieval Yes

4.5.2.26 Allow refinement of searches with qualifiers Yes

4.5.2.27 Ability to group: prior to search, user can delimit pending results' listings, and group results by item characteristics (type, format)

   Yes, it has option to sort results by Metadata field, order of the results and sorting by ranking.
4.5.2.28 Spell checker support  No

4.5.2.29 Refine searches: Once results are retrieved, users have the ability to focus and reduce number of items retrieved  Yes

### 4.5.3 Sorting

4.5.3.1 Author  Yes

4.5.3.2 Title  Yes

4.5.3.3 Issue Date  Yes

4.5.3.4 Relevance  Yes

4.5.4 Which search engine tool kit is used for searching the database?

CDS-Invenio uses Google like search engine tool kit for searching the database. It uses its own indexing technology and a search engine designed for large data repositories[4].

4.5.5 How is the recall and precision ratio while searching the documents?

Since the system supports searching via metadata fields and full text, the recall is high via word searching and precision is high if metadata fields are used for searching the database.

4.5.6 How is the speed of the retrieval of the documents?

The speed of the retrieval of the documents is very fast and CDS-Invenio shows number documents retrieved within number of seconds. This is one of the good feature of CDS-Invenio.

4.5.7 Does the system provide help for the users while searching digital documents?

Yes, proper help is provided on basic and advance searching. It also provides proper documentation for searching with examples.

4.5.8 How are the results displayed to the users? Whether users can download contents after searching? How are results displayed such as metadata, filetype, file size?
Users can download the contents after searching. Results are displayed along with all details and users can export data into different file formats such as BibTeX, Dublin Core, MARC, MARCXML, National Library of Medicine etc.

4.5.9 While viewing the results does the system provide access control? Yes

4.5.10 Does system support streaming audio/video option? Or it allows only to download documents? Whether browser can support any other document viewing? Does the end user need to have plugins and software for viewing particular documents?

CDS-Invenio supports streaming audio/video option. It can be integrated with existing streaming server. It also allows end user to download the audio/video. End user needs to have a required plugin installed for viewing a particular audio/video.

4.5.11 Does the system support advanced features such as my binders, virtual bag, shelving etc.

Yes, CDS-Invenio supports to store all search results in my basket option.

4.5.12 Whether system support unified search (Unified search is the support for finding items that are scattered among a distributed collection of information sources or services, typically involving sending queries to a number of servers and then merging the results to present an integrated consistent and coordinated format)?

No, CDS-Invenio does not support to search via unified search.

4.5.13 Whether system support federated search (federated search allow users to search multiple digital libraries with a single query from a single user interface. The user enters a search query in the portal interface’s search box and the query is sent to every other digital library software)?

Yes, CDS-Invenio supports to search via federated search.

4.6 Access Control, Privacy and Management

4.6.1 How users are managed in the software?

CDS-Invenio has an option called as "WebAcess Admin" where "Manage Accounts" option is available and where all details about each user are stored.
4.6.2 Does the system support to add/edit/delete user profiles? Yes

4.6.3 Does the system support to define different roles for different users?

Yes, one of the good feature of CDS-Invenio is it supports to define complex roles for different users.

4.6.4 Does the system keep detailed information of each user that registers in the repository such as name, email, phone number, address etc.?

CDS-Invenio keeps only name and email address of each user that registers into the repository.

4.6.5 How the permissions are given to users who have registered into the repository?

Repository administrator can define to give different permissions to different users who registers into the repository.

4.6.6 Does the system have provision to identify active and inactive members and take necessary action against inactive members? Yes

4.6.7 User Roles

4.6.7.1 Does system allow to create different groups for different users Yes

4.6.7.2 Does system allow to create different groups of users Yes

4.6.7.3 Does system allow definition of different user groups Yes

4.6.7.4 Does system limit access by role Yes

4.6.7.5 Does system allow collection to be customized for each role/group Yes

4.6.8 Does the system allow access limits at different levels?

(a) File/object level Yes

(b) Collection level Yes

(c) User/Group level Yes

4.6.9 How passwords are administered in the system?

CDS-Invenio keeps track of passwords submitted by users.
4.6.10 Does the system assign passwords for each user?

No, each user can select his/her own password.

4.6.11 Does the system allow users to select passwords? Yes

4.6.12 Does the system have mechanism to retrieve forgotten password? Yes

4.6.13 How is the level of access for Inside/outside users such as outside users can access only metadata and abstract level access and internal members can access full texts as well as metadata.

Access levels can be generated in CDS-Invenio at different levels such as collection/item level hence system administrator can determine which collection/item should be made public or not.

4.6.14 Does the system have any mechanism to provide access to the digital library via

(a) IP source address filtering Yes
(b) Proxy filtering No
(c) Credential-based access Yes

4.6.15 Are access denials flagged in any special manner by the software? Does the software keeps access logs and how accessible are the logs to either human or machine processing and interpretation?

Access denials are flagged in a special manner but repository does not keep track of access logs.

4.6.16 Does the system support access to the collection for a specific period? No

4.7 Authentication and Authorization

CDS-Invenio has multiple authentication methods that let the repository administrator define different levels of access and restrictions according to different roles performed by different users. Some documents might be accessible to anyone; others might be accessible to an academic community according to the IP addresses; and some others only to a
restricted number of users according to their IPs/passwords. CDS-Invenio restricts access according to the level of the File/Object. The repository administrator can restrict access to specific parts of a file. CDS-Invenio manages copyright issues for documents that have been submitted to it. For each specific collection maintaining different copyright policy is possible.

4.7.1 Does the system support to provide different authorization policies?  
Yes, CDS-Invenio supports to set different authorization policies.

4.7.2 System should support restricting access to contents through password authentication, IP filtering, and proxy filtering.  
CDS-Invenio supports to restrict access to contents through password authentication as well as can restrict access through IP filtering.

4.7.3 System should be able to support access based on browsing/viewing metadata, certain records, full text item or collections through different authorizations with different privileges for different communities. Yes

4.7.4 System should also be able to allow different roles for different members for different items/collections. Yes

4.7.5 Does the system supports to have a super user/admin user who is overall manager/administrator of the software? Yes

4.7.6 Does the system use authorization to support different roles within the system (for eg submit user, metadata editor, reviewer, approver, overall administrator etc.) Yes

4.7.7 Does the system use an external authentication mechanism (eg. Lightweight Directory Access Protocol). Yes

4.7.8 Does the system display only those pages that are pertinent to each user for which he/she is able to see? Yes

4.7.9 Does the system support the feature that groups are assigned to one or more roles? Yes

4.7.10 Does the system support the feature that users are assigned to one or more groups or collections? Yes
4.7.11 Does the system support the following roles:

(a) Content Developer Role: must be able to create/add/edit/delete metadata and content in defined (not ALL) collections in addition Yes
(b) Administrator Role: must be able to add/delete/create users and groups Yes
(c) Contributor Role: must be able to submit metadata and content, in addition to general users’ role Yes
(d) General User Yes
(e) Submitter Yes
(f) Metadata Reviewer Yes
(g) Max Role: must be able to do ANYTHING within the system Yes

4.7.12 Does the system use authorization to verify users with campus security systems (NetID)
Yes

4.8 Interoperability

4.8.1 It is necessary to know whether the software supports Z39.50 protocol and OAI-PMH protocol?
CDS-Invenio supports OAI-PMH2.0 but it does not support Z39.50

4.8.2 Which harvesting software is used by the Digital Library Software for metadata harvesting?
CDS-Invenio uses its own harvesting software.

4.8.3 SRU/SRW Support Yes

4.8.4 Does the software have OpenURL support?
Yes, CDS-Invenio supports openURL support.

4.9 Ease of Deployment of each software

4.9.1 Software Installation

4.9.1.1 Packaging and installation steps
Detailed installation process of CDS-Invenio is added in Appendix A of this document.

4.9.1.2 **Automatic installation script**

Software does not have any automatic installation scripts. The installation of the software is tricky one. One needs to have a good command on operating system for successful installation of the software.

4.9.1.3 **Disk space used**

700 MB disk space is required to install complete CDS-Invenio.

4.9.1.4 **Time required to install each software**

Installation of CDS-Invenio is very difficult. It takes more than a day to successfully run the installation if everything goes well.

4.9.1.5 **Documentation help available and useful for installation** Yes

4.9.1.6 **Environments needed to set for installation of each software**

No environments are needed to set for installation.

4.9.1.7 **Operating systems (Linux/Windows/Solaris/MacOS/GNU/AIX) on which software can be installed**

POSIX, Linux, Solaris, SunOS

4.9.1.8 **Programming languages used** Python, LISP and PL/SQL

4.9.1.9 **Database supported at back end** MySQL

4.9.1.10 **Web Server used** Apache Web Server

4.9.1.11 **Java servlet engine needed** No

4.9.2 **System Support/Maintenance**

4.9.2.1 **Documentation/manuals**

Online documentation of CDS-Invenio is available at [http://cds-invenio.cern.ch/invenio/documentation.html](http://cds-invenio.cern.ch/invenio/documentation.html)
4.9.2.2 Mailing lists/discussion forums

CDS-Invenio has three mailing lists such as

project-cds-invenio-announce@cern.ch
project-cds-invenio-users@cern.ch
project-cds-invenio-developers@cern.ch

4.9.2.3 Wiki pages

https://twiki.cern.ch/twiki/bin/view/cds/invenio

4.9.2.4 Help desk support  No

4.9.2.5 Ease of system administration (ability to configure for different uses)

CDS-Invenio is a complex software. It can perform different actions. It can be customized as per the end users requirements as well as it can define complex roles and for every role there are set of users and these users can perform different actions.

4.9.3 Hardware/Software Requirements

Because of the variety of possible conditions within each organization, specific demands on CPU cannot be predicted. System demands will depend on the size of the collections, the number of users, the database platform, etc.

4.9.3.1 What are the minimum hardware requirements for the repository software?

No specifications available on CDS-Invenio site.

4.9.3.2 How actively is the software maintained by its owners or contributors?

Though the first version of CDS-Invenio is yet to release but frequently there are latest versions of CDS-Invenio are available on web. CDS-Invenio version 0.92.1 was released on 20th February 2007 and now CDS-Invenio version 0.99.0 is released on 23rd March 2008.

4.9.3.3 Does the software have procedures in place to monitor and receive notifications when hardware technology changes are needed? No
4.9.3.4 Does the software creators have a process to stay current with the latest operating security fixes? Yes

4.9.4 Security

4.9.4.1 Does the system have security methods such as

(a) Data encryption No
(b) Digital signatures No

4.9.4.2 Server security (Does software has secure database connection?)
Yes, CDS-Invenio database is generated in MySQL hence it has a secure database connection.

4.9.4.3 Ability to restrict access repository at item level (For eg to view metadata but not content) Yes

4.9.5 System Level Security

4.9.5.1 Fixity - such as checksums or digests No

4.9.5.2 Management of user accounts and rights to specified locations within the DL No

4.9.5.3 Logging and auditing events
Since the software runs on Apache Web Server. All the error and even logs are kept in /opt/cds-invenio/var/log folder with apache.err and apache.log files. The installation script installs software in /opt/cds-invenio

4.9.5.4 Does the software support any security if OS doesn’t have any firewall implemented? No

4.9.6 General features related to technical aspect

4.9.6.1 Does the software have any back end maintenance? No

4.9.6.2 Does the software have RSS support? Yes
4.9.6.3 **Does the software have easy mechanisms to upgrade the software from old versions to new versions?**

Yes, CDS-Invenio migration is easy to carry out.

4.9.6.4 **Does the software support migration of data from one repository software to another repository software?**

Yes, CDS-Invenio supports to migrate data from other repositories into CDS-Invenio.

4.9.6.5 **Does the system support any ranking algorithms for the documents added in the repository?**

Yes, CDS-Invenio supports ranking of all documents that are added into the repository.

4.9.6.6 **Are there any known issues/bugs in the system?**

The current version has one bug and this bug can be rectified by using a small patch to make the software run successfully. The patch is called as websubmit_engine.py.

4.9.6.7 **Whether a novice user can easily publish content?**

Functionality of cds-invenio is difficult. One has to understand all functions properly then only end user can publish contents.

4.9.6.8 **How often user needs to use database schema for making any change?**

Users do not have to use database schema for making any changes.

4.9.6.9 **Whether user can reorganize database?** No

4.9.6.10 **Does system support automated content acquisition, harvesting and automatic metadata generation, including named entity recognition, automatic subject indexing/classification?**

No

4.9.6.11 **Does the system support to store metadata records separately from the actual contents?**

The metadata is not included directly with the full-text source. Metadata records are stored separately from the actual contents.

4.9.6.12 **Does the software support storing varied file formats?** Yes
4.9.6.13 Does the system keep the original file’s name, size and created date? Yes

4.9.6.14 While using the software if there are any errors occur does the system allow to submit the error report? Yes

4.9.6.15 How many developers are working for the software?

There are 2 developers working for the software.

4.9.6.16 How many users are using the software/strength of community?

There are only 19 installations so far of CDS-Invenio all over the world as on 20th April 2008.

4.9.6.17 Does the software provide its history on the web page for evidence of sustainability and vitality? Yes

4.9.6.18 Total number of downloads of the software?

No information about total number of downloads of CDS-Invenio available anywhere.

4.9.6.19 Is the software really under Open Source License terms and conditions?

CDS-Invenio is available under GNU Public license. CDS-Invenio distribution is free under the terms of GNU. There are no costs involved while using the software. It can be modified and redistributed for free. It is a freeware, but that does not mean that CERN undertakes any official obligation to provide help for its technical support and installation. If a library or other organization requires help, it can be provided for a charge that is defined after a special agreement. The copyright of CDS-Invenio belongs to CERN.

4.9.6.20 Are there any costs involved for using OSS-DL? No

4.9.6.21 Does the repository software have a clear policy regarding software upgrades? Yes

4.9.6.22 Whether the software has a long-term leadership and consulting services? Yes

4.9.6.23 After releasing the software whether it was tested and whether it has integrated new features or not?
The current version of CDS-Invenio 0.92.1 which is released on 27th February 2007 brought out lot of new features. So far the software has developed from 0.0.9 since 2002 to 0.92.1 in 2007. Though the development of the software is very slow it supports many advance functionalities. It has brought out different versions since 2002 though it has not yet reached to version 1.0 as on today.

4.9.6.24 How is the mailing list used by the users?

Though the software has a very good functionality it has not yet reached to large audience. There are around 19 installations of the software so far as the installation process is not that easy. The following table (Table 4.1) shows number of requests sent on CDS-Invenio user’s mailing list from 2002 as observed on 20th April 2008.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of requests sent on CDS-Invenio’s User’s Mailing List</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>13</td>
</tr>
<tr>
<td>2003</td>
<td>33</td>
</tr>
<tr>
<td>2004</td>
<td>147</td>
</tr>
<tr>
<td>2005</td>
<td>203</td>
</tr>
<tr>
<td>2006</td>
<td>87</td>
</tr>
<tr>
<td>2007</td>
<td>330</td>
</tr>
<tr>
<td>2008, April 20th</td>
<td>100</td>
</tr>
</tbody>
</table>

4.9.6.25 Whether software has conducted any workshops, training sessions on local, regional, national and international levels?

There is no mention so far of any workshop or training sessions held of CDS-Invenio.

4.9.6.26 Whether software is built according to open standards?

The software uses MARC 21, Dublin Core standards.
4.9.7 Architecture of the Software

4.9.7.1 How is the architecture and design of the software whether it is simple or complex?

Architecture of CDS-Invenio is not simple. It has a complex architecture. Data acquisition in CDS-Invenio is performed at different sources such as direct author submission, OAI and non-OAI harvesting. The key feature of CDSware’s architecture is modular logic where each module carries out a specific, defined functionality.[5]

![Figure 4.1: CDS-Invenio Architecture](image)

4.9.7.2 Does the architecture support separation between different local parts and put into different machines (e.g. centralised/distributed database, relational/object-oriented database management system, different components of the directory to distribute to different machines) and the transport model (protocols for communication between the system and the user interface or between system components). No

4.9.7.3 Does system support grid architecture? How grids can help to disaster recovery of data?

No information about grid architecture was found anywhere on CDS-Invenio site.

4.9.8 Scalability: Is necessary to know whether software is scalable in terms of storage and retrieval?

CDS-Invenio can handle millions of documents hence it is scalable in terms of storage as well as retrieval but above 4 million of records, it requires parallel architecture with multiple servers[6].
4.9.9 **Extensibility:** Extensible, to allow new elements (collections, data types, services, etc.) to be easily added to the digital library.

Yes, CDS-Invenio allows to add new collections, data types, etc. easily.

4.9.10 **Storage:** Does the system support data compression for e.g., in case of multimedia collections, the size of the files (for e.g., video/audio files) would be very large, hence the system should support compression/decompression techniques. Does the system support to store compressed file formats?

Yes, CDS-Invenio supports to store compressed file formats. Yes, CDS-Invenio also supports compression/decompression techniques.

4.9.11 **Backup/Restore facility**

4.9.11.1 Does software support backup function, which contains metadata associated with access controls, customized files, as well as full texts of the documents added in the repository?

CDS-Invenio does not provide any backup function.

4.9.11.2 Does the software explicitly require any particular backup strategy, or does it just rely on system-level backup plans, like periodic disk backups to tape?

One has to rely on system-level backup.

4.9.11.3 In the event of a disaster, what recovery plans are supported by the software?

There are no recovery plans in the event of a disaster.

4.10 **User Friendly Interface**

CDS-Invenio allows to modify the user interface through an API or through scripts, which control the presentation of this service. Headers and footers allow for the creation of static and dynamic pages. Each collection’s Home Page can be formatted according to their needs and desires. Each collection can have a separate interface. Multilingual interfaces are supported. CDS-Invenio supports user interface in 18 languages. CDS-Invenio allows users to save repository content in personal files inside the system.
4.10.1 Whether end user can easily customize the user interface "look and feel"? Whether end user can change the header, theme, footer, overall web interface of the system?

Yes, CDS-Invenio supports to easily customize the user interface "look and feel" as well as header, theme, footer etc.

4.10.2 Can user interfaces be localized in their language?

Yes, CDS-Invenio supports to customize the user interface in different local languages. At present user interface is available in Bulgarian, Catalan, Czech, German, Greek, English, Spanish, French, Italian, Japanese, Norwegian, Polish, Portuguese, Russian, Slovak, Swedish, and Ukrainian.

4.10.3 How is user interface for the experienced or non experienced users?

The terminologies which are used in CDS-Invenio are different hence it takes time for the Library staff to know all terminologies used in CDS-Invenio.

4.10.4 How is the user interface designed for the use of Librarian or for the use of users?

User Interface is same for the use of Librarian or for the use of user.

4.10.5 Whether system provide web based interface for all its functionalities?

Yes, CDS-Invenio carry out all functions through web user interface.

4.10.6 Does the system provide online help? Yes

4.10.7 Does the system provide multilingual access support? Yes

4.10.8 Does the system allow users to submit feedback on system problems? Yes

4.10.9 Does the system notify user’s current state of the task? No

4.11 Usability

4.11.1 What user skill levels are expected to use the web interface?

For users, user interface is easy to use for submitting documents or searching the database. But for the Librarian user interface is not very easy to use as well as there are complex functionalities.
4.11.2 Is the user able to navigate and find his/her ways to information quickly and easily? Yes

4.11.3 How are the help features provided with the software? Does help feature allow a user to feel comfortable within a digital library? Does the help feature provide general as well as technical answers to the problems? No

4.11.4 Does the user need to know which player or viewer is necessary to view multimedia or other digital objects from the system? No

4.11.5 Does the user interface provide facility to give user opinion on the web as well as does the software support user to fill up online user feedback form? Yes

4.11.6 Does the system support to show usage statistics such as usage patterns and which material is used more by which group, how often, when, for what reasons/decisions etc.:

(a) Usage patterns
(b) Use of materials
(c) Usage statistics
(d) Who uses what, when
(e) For what reasons/decisions

CDS-Invenio supports web usage analyzer based on general web site hits, web site hits per day, unique visitors per day, search collection usage analysis, search engine query analysis, user basket statistics, user alert statistics etc.

The unique feature of CDS-INVENIO is it generates ranking data for ranking search results based on methods like: Journal Impact Factor, Word Similarity/Similar Records, Combined Method. The software also supports "web comment" option where a user can put his comments against a particular record. Users can also put "reviews" for the documents they have browsed through.

4.11.7 Transaction Log Analysis

4.11.7.1 Does the system keep error log files? Yes

4.11.7.2 Does the system keep user logs? No

4.11.7.3 Does the system keep track of query's submitted to the system? Does it store any query log? Yes
4.12 Copyright/Policy Issues

4.12.1 Does the repository software have any means to manage, store, or enforce contract agreements for all the digital documents that are added into the repository? Yes

4.12.2 Does the software have any capabilities, such as access control lists, Internet address filters, etc., that can be used to enforce copyright or access restrictions? How granular are these access controls? Can different restrictions be applied to different objects in the repository? Yes

4.12.3 If repository ingests digital content with unclear ownership/rights, does it have policies addressing liability and challenges to those rights? Yes

4.13 Advanced Features

4.13.1 Whether software has long term leadership and consulting services? Yes

4.13.2 Does the system provide citation data? Yes

4.13.3 Does the software have defined road map for the future? Yes

4.13.4 Does software support indexing of all documents that are added in the repository in a distributed way, without central control? No

4.13.5 Virtual Collection Support: Does software supports to generate virtual collections /special collections across several content providers? Yes

4.13.6 Load Balancing: Does software has any mention of load balancing? It is necessary to know how the software functions when number of records are more or more number of people try to access the software?

CDS-Invenio can handle large volume of data. It has capability to handle millions of documents and perform well. If more number of people try to access the software it can handle those situations without any failures.

4.13.7 Visualization: Does the software support visualization feature? No
4.13.8 **Personalization: Does the system support Personalization feature?**

Yes, CDS-Invenio support personalization feature. Each user can create their own collection baskets and CDS-Invenio’s search engine alerts for the new items added in the repository to each user through email. User can also generate discussion on any document which he feels. User can also get citation information of each document. User’s can also write reviews of any documents that are there in the repository.

4.13.9 **Translation Service: Does the system provide translation service?**  No

4.13.10 **Community Services: Does the software support to provide community services such as allowing members of the digital library community to exchange ideas, make announcements, write reviews etc.**

Yes, CDS-Invenio supports to provide community services such as it allow members to exchange ideas, make announcements and write reviews as well as comments.

4.13.11 **Page Ranking: Does the software support page ranking such as supported by commercial digital libraries.**

Yes, CDS-Invenio supports page ranking algorithm for all documents that are added into the repository. Software allow end users to submit reviews about each document that is read by the user.

### 4.14 Digital Preservation

CDS-Invenio uses automatic conversion of any text file format to PDF format for long-term preservation. CDS-Invenio uses MARC 21 format to ensure portability and preservation. Software does not yet support several other features related to digital preservation. There is no mention of any digital preservation strategy for CDS-Invenio.

4.14.1 **Does the software support any digital preservation strategy and if yes, does it explicitly support any particular preservation strategy, such as described by PREMIS like, bit-level preservation, format migration, format normalization, emulation, or restrictions on submission formats?**

CDS-Invenio does not support any digital preservation strategy.
4.14.2 Whether system preserve file’s original identities such as its name, size and created date?

CDS-Invenio preserve file’s original identity such as its name and size.

4.14.3 Does the system have any data integrity check for a file added into the repository? No

4.14.4 Does the software have quality control measures to ensure integrity, and persistent documentation identification for migration purposes? No

4.14.5 Does the repository software preserve pre-existing persistent identifiers for submitted packages or objects? No

4.14.6 Does the software support preservation of metadata? Preservation metadata is information that supports and documents the long-term preservation of digital materials. Maintaining a history of a digital object with metadata is a key part of digital preservation strategy. It addresses the following attributes of an archived digital object:

- provenance - documenting the history of the object. Yes
- authenticity - validating that the digital object is in fact what it should be, and has not been altered. No
- preservation activity - documenting the actions taken to preserve the digital object. No
- technical environment - describing the technical requirements, such as hardware and software, needed to render and use the digital objects. Yes
- rights management - recording any binding intellectual property rights that may limit the repository’s ability to preserve and disseminate the digital object over time. No

4.14.7 How does the software manage compound objects (where multiple file formats of same object are linked together)?

CDS-Invenio supports to upload compound digital objects for e.g. it can upload postscript files and can automatically convert postscript files into pdf format. It also support to upload multiple file formats of the same document.

4.14.8 Does it have mechanism to keep licensing conditions for individual images/objects in the repository? Yes
4.14.9 Does the software support to track and manage copyrights and restrictions on use as required by contract or license? No

4.14.10 Where does the repository software store actual digital files and the metadata?
Actual digital files are stored in /opt/cds-invenion/var/data directory of CDS-Invenio installation and metadata is stored in mysql

4.14.11 Can the repository software automatically validate checksums on a periodic basis? No

4.14.12 Are the checksums cryptographically signed to prevent tampering? No

4.14.13 Does the repository software have a well documented process by which a submission is ingested into the repository for storage? For example, how and when does the software generate fixity data, such as checksums? No

4.14.14 How does the repository software verify that archival objects and metadata are correct? Are there automated checks of the metadata, such as to verify that a date entered into a field really is a date string? No

4.14.15 Does the repository software provide audit logs of all events that have occurred in the life cycle of a package? What events are logged?
Yes, CDS-Invenio records number of records submitted and errors occurred while submitting records, updated details etc. as a log data.

4.14.16 If repository ingests digital content with unclear ownership/rights, does software have policies addressing liability and challenges to those rights?
Yes, it is possible to address ownership rights to each digital object that is added into the repository.

4.14.17 Does the software have mechanisms to change its preservation plans as a result of its monitoring activities? No

4.14.18 Does the system have any mechanism in determining when objects in digital archives should migrate to new hardware and software? No

4.14.19 Can the software support scheduled events such that a human can be notified on a preset schedule to manually check for format obsolescence? No
4.14.20 Can new file formats be added or removed?

Yes, CDS-Invenio can add any type of documents into the repository. The software provide a feature to add different document types into the repository.

4.14.21 Does the software have ability to handle variety of file formats and does it also supports file format versioning?

CDS-Invenio can handle variety of file formats. It does not yet support file format versioning.

4.14.22 Does the software have mechanisms in place for monitoring and notification when format (or other representation information) obsolescence is near/or are no longer viable? No

4.14.23 Can the software monitor any standard format registries in order to ascertain format obsolescence? No

4.14.24 What standards does the repository software use to describe file formats and does the software record representation information? Does it use Internet MIME Types? No

4.14.25 Does it use format registries? If yes, which format registry is used such as PRONOM or GDFR or DCC to represent format information? No

4.14.26 Does the system support automatic format registration, for unknown formats? No
Figure 4.2: CDS-Invenio Installation Screen
Bibliography


