

GENERAL SYSTEM SPECIFICATION

Digital Collections **DC-X**

Baseline Version 1.6—Rev. 12/7/10

DC-X Digital Asset Management System

General System Specification

Baseline 1.6

Overview

The DC-X digital asset management system is specially designed to meet the functional requirements of an organization—for managing all of its digital content, e.g., for print and electronic publishing (web and/or mobile).

The DC-X system consists of software components from Digital Collections GmbH and GMTI, and several state-of-the-art third party (Open Source) software components including, but not limited to, Apache web server, MySQL database, and the PHP scripting language. Linux is the supported Operating System. Certified Linux distributions include RedHat / CentOS, SuSE Enterprise Linux and Ubuntu Server.

In the context of “Software as a Service”, GMTI will provide all application software, and other third party software components as required.

The following specifications pertain to the Baseline 1.6 commercial version of the DC-X system.

1. DC-X Technology

DC-X is a digital asset management system based on Internet/Intranet technology, relational and object databases, and client-server system architecture. The database has the capability to index a wide variety of multimedia objects, text records and native file types for storage and retrieval.

The server software consists of the DC-X client application running on a Linux[®] operating system and the MySQL database (with the Solr/Lucene search engine for full-text queries and information retrieval). The Oracle database is optional.

Open source utilities are employed, including Apache, PHP, Ghostscript, ImageMagick, FFmpeg and more than three dozen others.

Also included at the server level is the Digital Collections application programmer interface (API). The API provides an independent front-end application server which creates a rapid deployment environment for modifying the graphical user interface (GUI). There is no programming overlap between the user interface (the application layer) and the server platform technology underpinnings (the system layer).

The system software embraces the REST (*Representational State Transfer*) architecture style, and makes extensive use of the *AtomPub* HTTP-based protocol for creating and updating web resources.

The DC-X database file server exists as a WWW site. All functionality in the database application is controlled and performed through standard Internet browsers such as Mozilla/Firefox, Safari, and Microsoft Internet Explorer.

The DC-X system database enables simultaneous storage and retrieval of text records, native file types and multimedia objects.

Access restriction for operating system administration and database administration is handled through the use of the DC-X System Administration application—through a common web browser. Individual and group user accounts are set-up and maintained through the use of the system administration utility. These accounts cannot be used for system logins at the operating system level.

The DC-X database can be configured with multiple exclusive data pools (“Channels”) for different user groups or individual users (or, publications, etc.). Access to each Channel is restricted to privileged users (i.e., users must have the privilege in their user profile to access a given Channel). The database design and implementation can be tailored to fit special buyer requirements.

2. Server Hardware Technical Specifications

Linux Server

Minimum: Intel or AMD quad-core processor, 8 Gigabytes RAM

Internal or External Storage Sub-subsystems (RAID)

Specifications vary depending on buyer requirements. GMTI provides specifications at buyer’s request.

3. Relational Database Application & Redundancy

The relational database application is MySQL with the Apache Solr/Lucene search engine tightly integrated with the MySQL database server. The Solr/Lucene search engine enables state of the art full-text search and retrieval.

The DC-X database has the capability to maintain relationships between assets for indexing and searching. For example, if you perform a search for photos of George W. Bush, and find the one you’re looking for, you might also like to then retrieve any story that may have run with that photo—or vice versa. Another example would be a PDF page: if you find a (PDF) page of interest, you might like to in-turn retrieve all the stories and/or photos that ran on that page. When retrieving an individual digital asset, any files linked to that asset are displayed for access via a mouse click.

Along with this, a DC-X user can concatenate separate asset files into a single file (i.e., a photo and the story that it ran with, or vice-versa).

DUAL SERVER REDUNDANCY CONFIGURATION: The baseline DC-X System includes redundancy software that can be configured to support a hot backup configuration—providing a real-time copy of the master database that can serve as a fail-over system. In the unlikely event of a fatal system error on the master system, the slave (hot backup) system becomes the new master system.

4. Digital Rights Management

Digital rights management (DRM) is supported through the use of rights-specific metadata fields and a variety of underlying processes that act on those fields. Information that is not tagged as qualifying for public access is prohibited from being indexed in any public access areas of the DC-X database (or for commercial distribution). In general, DC-X developers recognize the Picture Licensing Universal System (PLUS) as a basis for managing digital rights as they pertain to images. An explanation of PLUS can be reviewed at the following URL: <http://www.useplus.com/aboutplus/system.asp>.

5. Backup strategy

All digital content is stored outside the database. Online backup of database contents can be done automatically.

Restrictions apply regarding backups such that only a full restore to a database state is possible. Partial restore of backups are not supported due to database design—because, in this case the referential integrity cannot be maintained.

Operating system backups and application software backups are done on an as-needed basis (e.g., when a software modification is made or a version upgrade is done), since these software components do not change every day.

6. DC-X Client Application Functional Specifications

The DC-X client application runs within a common web browser such as Mozilla/Firefox, Safari, or Microsoft Internet Explorer.

Since the DC-X graphical user interface (GUI) is within a web browser, no special requirements exist for desktop workstations. A reasonably configured computer running Windows® or Mac OS® is sufficient. The minimum required screen resolution display setting is 1024x768 pixels.

At any time, third party applications may run on the same workstation, such as spreadsheet software or desktop publishing applications.

Calculators and notes windows are part of the workstation operating system, and may also be used simultaneously with the DC-X client application.

At buyer's discretion, email service can (optionally) be implemented for the DC-X server in order to support email communications for all DC-X client users—driven by the DC-X system, or for sharing or distributing content.

The DC-X baseline 1.6 client application consists of three modules. Each module is represented by a labeled tab—accessible from the (horizontal) parent menu bar. Following is a brief description of those three modules.

6.1) News Cockpit

The News Cockpit is a dashboard for news monitoring, research and data acquisition. It includes a search mask for simple and advanced full-text search, and/or using tag-based search criteria such as keyword, category or classification (or any other metadata). For customers who purchase the DC semantic engine (Elektra) web service, semantic analysis and filtering facilitates a “find” versus “search” approach to content research and acquisition. “Related Content” and “External Content” are tabbed functions that enable the user to find content in the DC-X database that is related to an item (or items) in one's search results, or, conversely, similar content that is external to the DC-X database but indexed by the semantic analysis of existing RSS data feeds. Tools for the research, gathering and sharing of content also include content “Agents” (search agents, as in saved searches that post query results with a single mouse click), and “Tags” (public and private collections of user-selected content). All database indices (Channels) can be accessed and queried via the News Cockpit module—for both internal and external content. The News Cockpit is the main module for enterprise content sharing.

6.2) Story Editor (optional software)

The Story Editor exists for content production and packaging, and includes a full set of authoring tools for writing, editing, and packaging content for the online, mobile and print publishing channels. The integrated Story Editor features the automated acquisition of “related content”, i.e., content that is related to the story one is in the process of writing. Related content is displayed in a separate adjacent frame—for quick and immediate reference.

6.3) Administration (DC-X System Administration Utilities)

The Administration module includes four menu options:

- 1 - Users and Groups: user account setup and administration
- 2 - Workflows: hot folders, data feeds (imports/exports, AtomPub/RSS feeds and statistics)
- 3 - Content: channels, event flags, field definitions, topics mapping, rights mapping
- 4 - Configuration: all (system configuration) settings, language translations

7. The DC-X Basic System

An entry level offering consists of a basic DC-X System. It provides the basic functions of a DC asset management system and is based on the MySQL database. The DC-X Basic System includes the following components:

- **Asset Manager** – The Asset Manager is the basic module of DC-X. It provides the basic functions of the system and includes an easy-to-use system administration utility for managing the entire system. The system administration utility is accessed via a web browser. All digital content is stored in one central pool. Import of documents (see input filters listed below). Basic functionality includes: additional keywording of meta-data (for improved search results), high performance full-text search based on the Lucene full-text search engine, management of documents, exports via download, and copy-and-paste of content.

NOTE: Geo-Tagging Support

Google Maps/Google Maps-API: a Google Maps-API is required in order to implement Google Maps in DC-X. Depending on the terms of use of Google, the customer must pay for the Google Maps-API. The costs of the Google Maps-API Premier is not included in any DC price quote, however; the support for using it is included with the basic DC-X system. For more information please visit the link: <http://code.google.com/intl/en/apis/maps/index.html>.

- **DC-X Uploader** - The DC-X Uploader is an Adobe Flex based upload client that integrates tightly with the desktop. By simply dragging and dropping files from your desktop, Apple Finder, Windows Explorer, or from one of the many other supported clients—into the uploader, the user can import data into DC-X. The DC-X uploader also provides a simple tagging functionality with optional as well as mandatory fields; e.g., caption, title, keywords.
- **Content Agents** - Content Agents execute searches automatically and notify the user based on his/her profile. Distribution of Content Agents, i.e., search results, is provided by the Content Agents.
- **AtomPub Interface** – The AtomPub interface can be used to remotely access DC-X with third-party applications.
- **API: Semantic Engine (Content Tagging and Recommendation)** - Integration with DC's Content Tagging and Recommendation/Semantic Engine provides automatic tagging and categorization of documents. Personalities, Organizations, Cities and Countries will be recognized and added as metadata.
- **API: OpenCalais** – Integration with the (Thomson Reuters) OpenCalais webservice. An OpenCalais account is required (<http://www.opencalais.com>), which might involve additional expenses based on the monthly usage. DC can not guarantee the availability of the service, since it is not provided by DC.
- **Picture** - Import of images (JPEG and TIFF); includes the extraction of IPTC, EXIF and XMP information.
- **Video** - Import of video files (MPEG-2 and MPEG-4). Extraction of meta-data, e.g., frame-rate, length etc.
- **PDF** - Import of PDF documents. Extraction of full-text information for indexing and creation of previews (first page of document).
- **Native** - Import of all other file formats, including full-text indexing and generation of previews,

where supported. Depending of the versions the following file formats are supported: MS Word, MS Excel, MS Powerpoint, OpenOffice, PNG, BMP, RTF, EPS. Please ask about other file formats if not listed here. Additionally, the native scanner supports the import of any other file format, if metadata and preview file are provided.

- **DC-X Page Importer** - Import of pages from other systems, based on the DC-X standard.
- **Plug-In Framework** - With its Plug-In Framework, the DC-X application can be enhanced through the creation of program extensions using the PHP scripting language. The DC-X application layer is extendable without touching the DC-X core system layer. A Plug-In can be created to execute whenever a specified event occurs; e.g., new hits for a search agent are found, or the disposition of a document changes (e.g., to *published*, or, *ready for publication*).
- **DC Sync** - This is an application-based tool to synchronize two DC-X servers, e.g., the complete data or only data that has been manually selected by users. Utilizing DC Sync, the customer cannot only synchronize the entire system to a disaster recovery site for fail-over, but can also provide/sync content to an external web site that has a public-facing wall.

8. Fault Tolerance and Redundancy

Within the baseline 1.6 version of the DC-X system, database redundancy is maintained via the DC Sync. In a database redundancy configuration, the MySQL database is distributed to two servers. All transactions going on will be replicated across both servers, thus maintaining two like copies of the DC-X database.

In a dual server configuration, if the master server fails, the backup server will then take over the master sever roll and begin receiving and storing all data. Either of the two servers can be taken out of the network configuration for maintenance without interrupting system processing occurring on the remaining database file server.

Dual servers each have the complete suite of DC-X application software installed. Users simply have to choose a different WWW address (i.e., bookmark) if the server they are working on fails (in order to continue their work). In the event the master server fails, all saved transactions are available on the surviving (backup) server, and are available to client users once they switch-over their log-in session to that server.

9. Administration and Technical Support

Remote access to any DC-X system component can be granted to system administrators.

Since the system is implemented through a cooperative project effort (i.e., between buyer and seller), it is anticipated that buyer's personnel will be actively involved throughout the entire installation project, and will achieve the necessary level of technical proficiency for management of the system. A support contract is offered as part of GMTI's Purchase and Licensing Agreement. GMTI offers hot-line telephone technical support 24-hours-per-day, 7-days-a-week (including holidays).

10. System Response Time

As a goal, a query response time of 6 seconds is the standard for the first result set. Non-standard or very complex queries may exceed this time frame—depending on the database setup, storage system and available network band-width.

11. Other Standard Software

Print spooling will be done using the server operating system facilities, as well as any network print servers—such as output management systems.

12. News Wire Agency Reception Modules

DC-X system contains optional software to receive news wire transmissions, or “feeds”. This software is extended during the project to accommodate additional buyer requirements for news wire capture and management—for both photos and text—depending on buyer requirements. The basic DC-X System includes one news wire channel input. Additional channels are available at an additional cost.

Wire transmissions are stored in the database, maintaining all the original information received in the initial transmission. Automated processes will purge those digital files from the database after defined intervals—specified for each wire service, and, depending on specific buyer requirements.

Editors may receive copies of original news wire transmissions—as desired—for their own purposes. The ability to change original news wire takes is prohibited by design.

13. Data Filters and Legacy System Interfaces

Software modules are available for accommodating data transfer requirements, e.g., from legacy systems to the DC-X system. When necessary, such filters are custom-developed by GMTI to meet buyer specifications. One legacy archive database data filter and conversion is included with the base DC-X system. Data filters and conversions for additional legacy databases are available at additional cost.

Thumbnail previews

DC-X data filters have the general capability to automatically extract preview thumbnails from any files that contain them at the time they are processed by a DC-X data filter. Thumbnail previews are indexed and linked along with the original file and its contents. The thumbnail is then displayed in three formats: contact sheet (postage stamp size), light-box, and poster sizes, along with the file’s associated metadata. This includes PDF files.

14. External Access

External access to the DC-X system with restricted or non-restricted privileges is enabled using Internet technology.

15. Report Generation

Informational reports may be custom-designed—as a system option—and implemented in accordance with buyer specifications.

16. Workflow Description

Document Editing

DC-X system users research documents in the DC-X database. Documents can be viewed partially or completely. A user scrolls through a document without loading additional buffers from the server—using a

word processor-like HTML editor when editing is required. A “dossier” view of a document is also provided as a “read only” function.

Users collect documents of interest into a “Tag Group” (collection). There is no limitation on the number of Tag Groups for an individual user or user group.

Text editing is performed through the use of the DC-X HTML editor. The text editor is accessed via a single mouse click, and is displayed within the web browser. Users can view text in normal, bold, italic, underscore and notes mode for editing. Spell-checking is available, and is done on demand. The text editor supports a find-and-replace function as well as a “re-do/un-do” function. The DC-X text editor has basic “word-processing-like” capabilities.

Documents retrieved from the database are guaranteed to be free of unwanted invisible characters. Unintentionally entered characters of such nature will be stripped during a “save” transaction.

Users can review their work in the database and commit changes or assign deletions to their work. Deletion privilege is only available for authorized users. Deleted documents will be purged from the system automatically after a defined dormancy period. Within this dormancy period, authorized users may “un-delete” items that have been flagged for deletion.

All document entries in the database are search-able by any criteria; full-text, metadata/tags and combinations thereof.

Since all metadata is stored in the MySQL database, all documents and versions thereof will have unique system-generated identifiers. Stories and other database objects may be downloaded to a local DC-X user’s workstation and manually renamed in the process of doing so.

Content Handling

Capabilities for handling digital content include, but are not limited to:

- Ingestion and collection of content from a wide array of sources, including internal and external systems/sites, email and social media
- Semantic analysis and categorization of content, allowing for automated sourcing and filtering of text content
- Extraction of metadata and statistics from various content types including text, photos, various media and application files
- Workflow automation that can trigger various functions, including export or publication, based on story or content disposition, or flags—to online/print/mobile channels—exclusively or inclusively
- A flexible publication rights management system (digital rights management; copyright ownership)

17. RSS Channel Feeds and Configuration

The DC-X system for collecting and importing RSS content is entirely configurable using the standard administrative utility. It’s as easy as creating a new entry, specifying where to look and how often to check for new material. It’s also possible to alter other components like DC-X channels to route a set of feed contents to a particular content pool or group of users, and to attach tags or flags to the content as it is imported.

18. DC-X Baseline 1.6 Database Specifications

The DC-X multimedia database consists of the MySQL server for Linux. The baseline 1.6 database includes support for the indexing, storage and full-text search/retrieval of the following file types (and many others too numerous to mention here):

- 18.1 XML and variants thereof
- 18.2 NITF, NewsML and variants thereof
- 18.3 JPEG RGB/TIFF
- 18.4 GIF
- 18.5 Vector Graphics: Adobe Illustrator®
- 18.6 Portable Document Format (PDF) files—including text content
- 18.7 QuarkXPress files
- 18.8 Adobe InDesign files
- 18.9 MS Office Files, including MS Word and MS Excel
- 18.10 MPEG, FLV, MOV and other mainstream video formats
- 18.11 WAV and other mainstream audio file formats
- 18.12 HTML, DHTML
- 18.13 PNG (portable network graphics)
- 18.14 Metadata associated with any of the above file types

19. DC-X Optional Software Modules (not included with the basic system)

- a) **Commercial/agency wire feeds:** Additional data feeds such as commercial news wire channels like Reuters, Bloomberg, etc. *Note: one commercial data (agency) feed channel is included with the basic system (typically a news photo service input like AP Photos).*
- b) **Production systems:** Interface for importing and/or exporting data from/to a production system. Note: two production system interfaces are included with the basic system: specifically, an interface to a print production system and as well as an interface to a content management system (for online publishing). Additional interfaces are available for additional costs.
- c) **E-Mail:** Import of E-Mails and extraction of attachments. This option supports POP3 and IMAP. The mail server is not part of the basic system and must be provided by the customer. Customers can create a black-list or white-list to exclude or included specific content providers.
- d) **RSS/Atom feed importer:** Import of RSS and Atom feeds (automatically downloaded, enclosures downloaded); extraction and indexing of full-text information.
- e) **IntelliTune interface:** Export of photos including processing instructions (crop, rotate, mirror, scaling to target resolution) to a hot folder managed by IntelliTune, and re-importing them after processing. Re-imported images will be displayed in a widget inside DC-X. This option does not include an IntelliTune license.
- f) **Rules Engine:** The Rules Engine enables the administrator to define rules to modify metadata based on given metadata. Rules can be defined in a description language similar to "IF", "THEN", "ELSE"—in common programming languages; e.g., "IF the picture has the status 'can only be published once', THEN set status to 'blocked'".
- g) **Digital Rights Management:** Site-specific requirements to be specified prior to the installation project start date (subject to the terms of the contract agreement)

- h) **News distribution:** Site-specific requirements to be specified prior to the installation project start date (e.g., Amazon/Kindle, ProQuest, LexisNexis, etc.); subject to the terms of the contract agreement.
- i) **Kerberos Single Sign-on:** Utilizing the Kerberos protocol, DC-X provides an automatic authentication after the user has logged in on the workstation. Up-front, it needs to be clarified, if the combination of browser, operating system and authentication service must be supported (NTLM, Authentication).
- j) **Intelligent Storage Manager:** The DC-X Intelligent Storage Manager provides configurable storage management to distribute files over multiple storage devices. With this add-on, the customer is able to store frequently accessed data (e.g. low-res images, recent agency content) on faster and more expensive devices, and occasionally accessed data (i.e., high-res images, archive content) on slower and less expensive devices. Depending on the amount of data, this may lead to a significant reduction of storage costs.
- k) **DC External Application Server:** Security is most important. Therefore, it is important to prevent unwelcomed external access to your DC-X production system. The external application server is supplied with data from the internal application server, and makes the data available for access by external third party applications, utilizing the DC-X AtomPub interface.
- l) **DC-X Crowd Sourcing:** A comprehensive application module for crowd sourcing activities will be available as an optional software module in Q4-2011.

20. DC-X Publishing Suite (optional; not included with the basic system)

The following modules can be licensed in addition to the DC-X Basic System:

- **Story Editor** - Content Creation for recording content using the WYSIWYG editor. Includes the "Content Sharing" and "Content Recommendation" functions.
- **Basic Element Drag & Drop** for InDesign-based editorial systems - A Drag-and-Drop icon is displayed next to images and texts and can be dragged and dropped into editorial systems. The icon provides a URL pointing to an XML file containing all information regarding the dropped item; e.g., metadata, URL and/or associated directory path pointing to where the high resolution image resides.

21. Mobile Apps for DC-X (Optional; not included with the basic system)

A mobile application is available for accessing the DC-X system using Apple's iPad and other tablets.

22. Web Services (Optional; not included with the basic system)

Content Tagging and Recommendation (DC Semantic Engine): The DC Semantic Engine web service provides automatic detection of subjects, persons, organizations, events and geographical aspects (countries and cities). On the basis of approximately 500 million fragments of knowledge, the Semantic Engine analyzes digital information. Complex heuristic, statistical, semantic and linguistic methods analyze the essential characteristics of the content of a document and assign the significant thesaurus terms and keywords to that document. This, in turn, allows for the "content recommendation" (of related multilingual content). For each asset that is located in the system, DC-X automatically recommends related content. For example, when a user views a picture received from the Reuters news

agency, DC-X automatically displays related content from company-internal sources, or external sources such as other news agencies, news portals, or user-generated content sources like Flickr and YouTube.