6.3

Luna Imaging Inc.

Insight®

Insight 6.3
Command Line Importer

Insight®

This document contains information proprietary to Luna Imaging, Incorporated (Luna). This document is provided to you as part of the Luna Imaging, Inc. licensed software and is governed under the same licensing restrictions that apply to that license. You may reproduce this documentation, which contains Luna Imaging, Inc. proprietary information for use within your organization. You may not disclose or distribute this documentation to third parties.

Even though Luna has tested the hardware and software and reviewed the documentation, Luna makes no warranty or representation, either express or implied, with respect to the hardware, software, or documentation, their quality, performance, merchantability, or fitness for a particular purpose. Luna has made every effort to keep the information in this manual current and accurate as of the date of publication or revision. However, Luna does not guarantee or imply that this document is error free or accurate with regard to any particular specification.

In no event will Luna be liable for direct, indirect, special, incidental, or consequential damages resulting from any defect in the hardware, software, or documentation, even if advised of the possibility of such damages. In particular, Luna shall have no liability for any programs or data stored in or used with Luna products, including the costs of recovering such programs or data.

No Luna agent, dealer, or employee is authorized to make any modification, extension, or addition to the above statements.

All copyrights and trademarks belong to their respective holders. Insight and Inscribe are registered trademarks of Luna Imaging Inc.

TABLE OF CONTENTS

Insight Command Line Importer

This document is intended mainly for Insight system administrators to provide context and information on how to configure the Insight command line importer. Readers are expected to have a basic knowledge of the UNIX / Windows Operating Systems as well as knowledge of XML documents.
Insight Administrators should have a thorough knowledge of the Insight suite of tools and how to manage and administer Insight collections.

Figure 1: Insight and LUNA Component Diagram
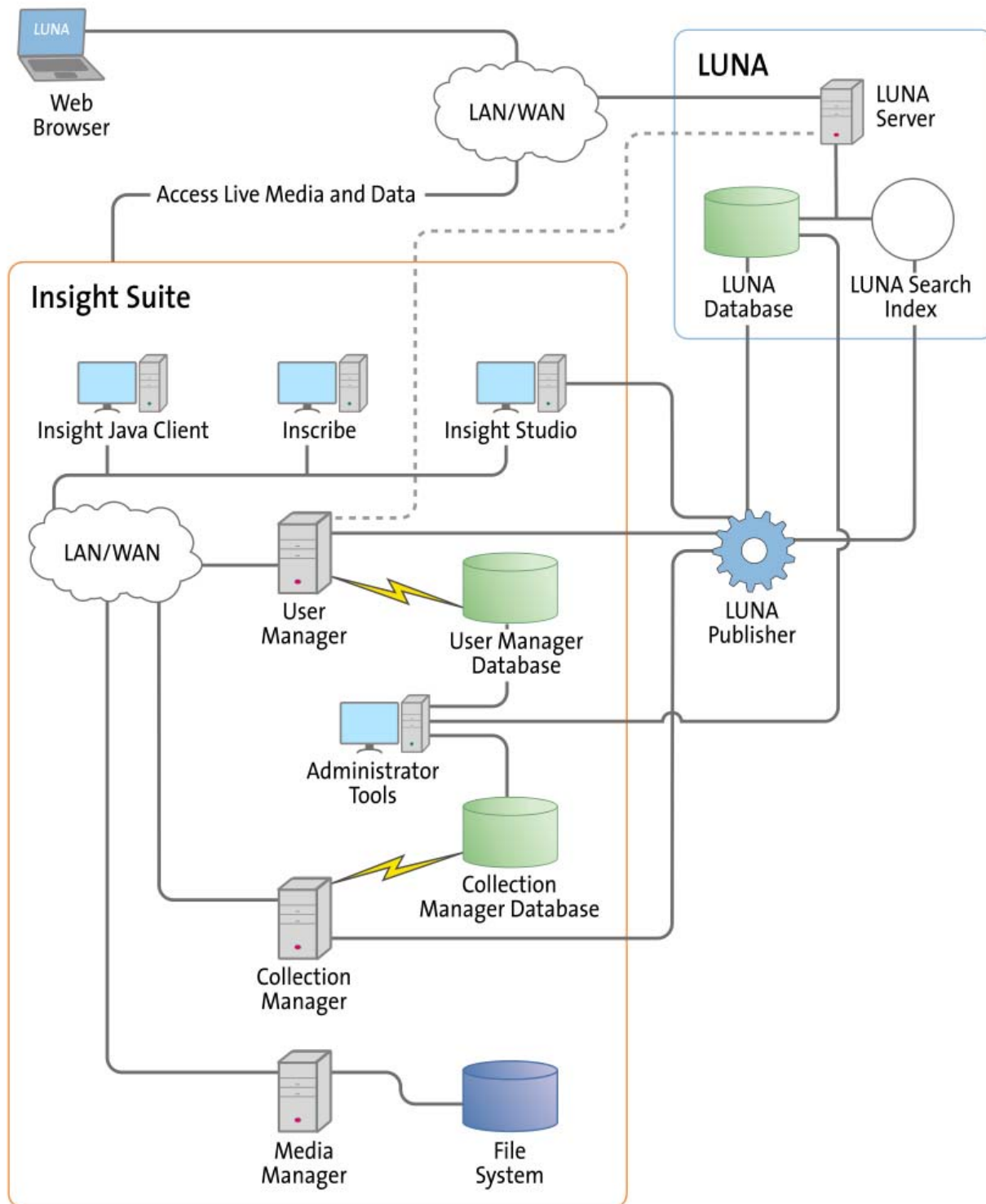
LUNA Components are as follows:

LUNA Publisher: Schedules and is used to pull data out of the Insight Collection Manager, storing it in an XML format. The LUNA Publisher then converts the archive data into a Lucene index for each collection and merges all Collection indexes into one single search index. Once merged, the LUNA Server is placed into a maintenance mode and the newly merged index replaces the existing one.

LUNA: LUNA is the dynamic web based front end to Insight collections. With LUNA, users can perform simple keyword searches as well as more complex structured queries, and control how they view, browse, and organize their results to create groups and presentations. LUNA includes the Web 2.0 tools your users are demanding which enable them to save and share their work by linking and embedding any view, image, group, or presentation.

LUNA Server: The LUNA Server is the server side component behind the LUNA client. Collections published to LUNA are indexed on the LUNA Server. The LUNA Publisher will automatically update the indexes based upon a schedule that you define.

Insight's Components are as follows:

User Manager: The User Manager functions as a single point of entry for all Insight Collections and Personal Insight Managers. It consolidates the functions of authentication and authorization for all shared resources within Insight and for LUNA. The User Manager can integrate with an existing security infrastructure if an institution already has a single sign-on solution in place. The User Manager also provides access to resources such as shared folders and groups within Insight. LUNA enables end-user management of viewing preferences, and storage folders for media groups and presentations.

Collection Manager: The Insight Collection Manager provides a common interface between client requests and the underlying data repository. Each Collection Manager may contain multiple collections and Virtual Collections of consistent or heterogeneous structures. The Collection Manager also functions as a broker between the clients (the Insight Java Client, Inscribe data editor, Studio (LUNA Publisher and XML Gateway) and the underlying database, coordinating search requests, and data updates.

Personal Insight Manager: The Personal Insight Manager is a specialized version of the Collection Manager which supports the creation of Personal Collections with the Insight Java Client.

Media Manager: The Insight Media Manager is built around a basic JSP Server, and manages access to Insight's media content. The Media Manager supports direct upload of processed content and also manages access to the JPEG2000 wavelet images that power Insight's Image Workspace.

Java Client: The Insight Java Client provides an additional interface for searching, viewing, and managing images, audio, video, and other media. The Java Client also provides end-user collection building tools with Personal Insight and the ability to create Virtual Collections – a subset of content with different access levels.

Inscribe®: The optional Inscribe data editor is an end-user cataloging tool designed to support all of an institution's cataloging needs. The Inscribe data editor includes built-in support for controlled vocabularies, date and numeric validation, and complex data relationships. Inscribe also supports an institution's workflow requirements by facilitating the publishing and review processes.

Studio: Studio provides collection administrators with the tools to build and manage collections. Studio is a wizard-based tool supporting batch processing of media, batch import of metadata, user management, catalog template creation and management, and other basic functions. The LUNA Publisher is accessible from within the Studio client.

Administrator Tools: The Insight Administrator Tools allow a collection administrator to dynamically manage collections. Components of the Insight Administrator Tools include tools for batch media processing, user and resource management, data indexing, and configuring of the search and user interface.

JPEG2000 Wavelet Encoder: The Administrator Tools and Studio include a JPEG2000 Wavelet Encoder. Wavelet encoded images improve performance when accessing large images, reduce storage requirements, and optimize network usage on image delivery.

Insight XML Gateway: The XML Gateway provides a web-services based interface for searching and retrieving content stored within Insight collections. The XML Gateway is middleware which allows backend interaction between Insight and other XML aware applications.

Introduction to Insight Command Line Importer

The Insight command line importer will allow administrators to load images and data into an Insight collection from the command line without the need to use Insight Studio.

Data records can be loaded alone or you can load both data records and their associated media.

This import method requires your data records to be in Insight's XML format for your specific Catalog Template.

If you will be importing media along with your data records, you will need to add a media section to your XML document as described later in this document.

Installing the Insight Command Line Importer

Preinstall

You will need to know the following before running the installer. You can locate this information using the User Manager/Collections section of the Administrator tools.

- Your Insight User Managers Address
- A studio user and password account with access to the Collection you want to add content to.
- The Collection ID for the Collection
- The Institution ID for the Collection
- Virtual Collection ID (normally NA)
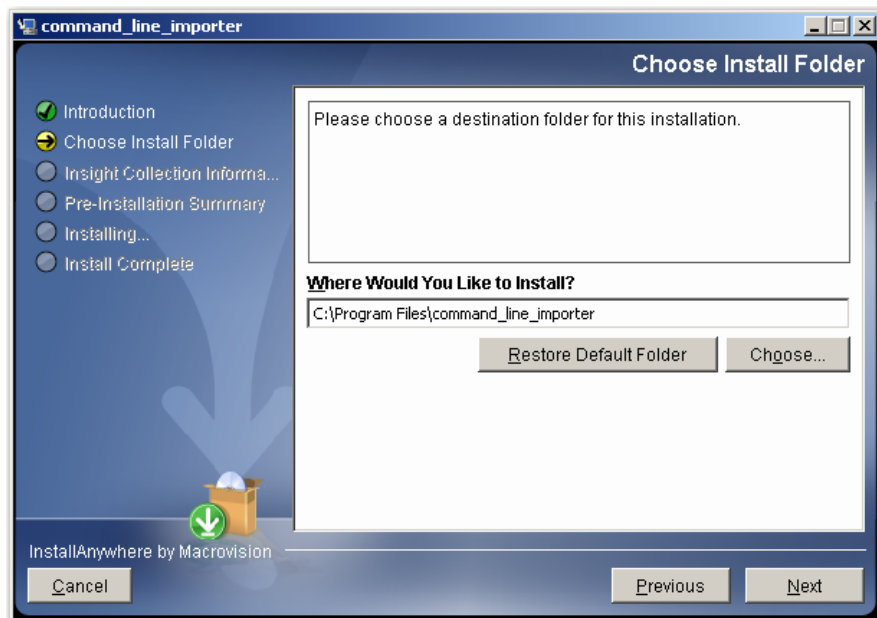- The Linking field used to link the image to the data record

Install

Locate the installer on CD1 in the utilities folder called cmd_line_importer.

Launch the Installer and follow the instructions in the installer.

To run the Installer without a GUI, on the command line use the following argument:

```
CLI_install –i console
```

Choose your Install location.

Fill in the Insight Collection information that will be used to configure the import.properties file. This is the information that will allow the utility to load content into a specific collection.
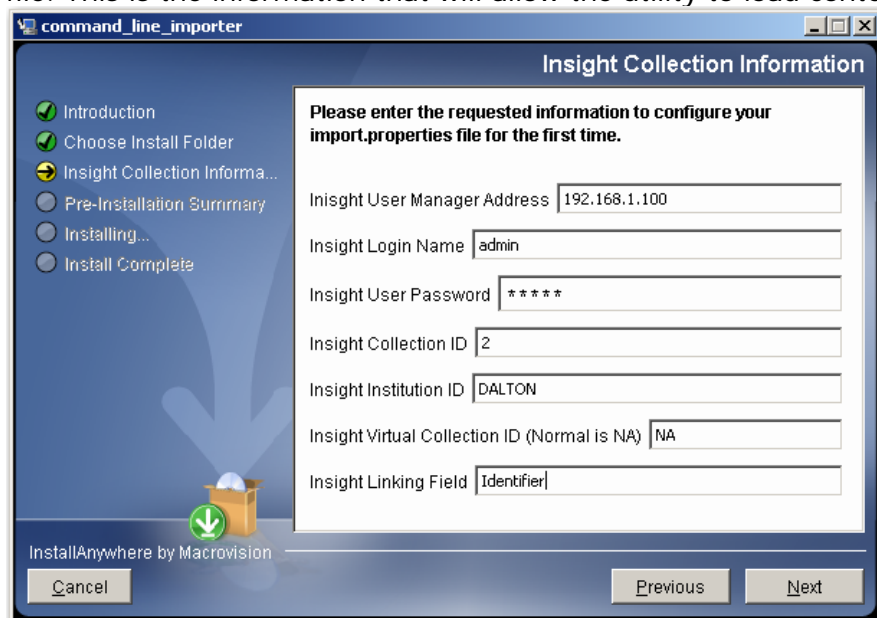
Using Insight Command Line Importer

Configuration

The import.properties file is located in the root of the install directory.

The only section of this document you need to configure is the section beginning with "import." All these setting were defined during the install.

```
import.userManager=192.168.1.100
import.username=admin
import.password=admin
import.cid=2
import.iid=DALTON
import.vcid=NA

import.linkField=Identifier
import.batchSize = 25

#import.logdir=S:\\work\\Importer\\log
#import.imagedir=S:\\work\\Importer\\images\\

#spider.database.connection = jdbc:mysql://localhost/spider
#spider.database.username = myuser
#spider.database.password = mypass
#spider.database.classname = com.mysql.jdbc.Driver

########################### log4j ####################
log4j.rootLogger=DEBUG,stdout, F, err
log4j.appender.F=org.apache.log4j.FileAppender
log4j.appender.F.File=import.log
log4j.appender.F.layout=org.apache.log4j.PatternLayout
log4j.appender.F.layout.ConversionPattern=%d{dd/MM/yy HH:mm:ss} | %-10t | %C{1} (%L) %x -
%m%n
log4j.appender.err=org.apache.log4j.FileAppender
log4j.appender.err.File=import.error.log
log4j.appender.err.layout=org.apache.log4j.PatternLayout
log4j.appender.err.layout.ConversionPattern=%d{dd/MM/yy HH:mm:ss} | %-10t | %C{1} (%L) %x -
%m%n
log4j.appender.err.Threshold=ERROR
log4j.appender.stdout=org.apache.log4j.ConsoleAppender
log4j.appender.stdout.layout=org.apache.log4j.PatternLayout
log4j.appender.stdout.layout.ConversionPattern=%d{dd/MM/yy HH:mm:ss} | %-10t | %-20C{1} (%L)
%x - %m%n
log4j.logger.org.apache=INFO
log4j.logger.com.luna=INFO
log4j.logger.httpclient.wire=INFO
```
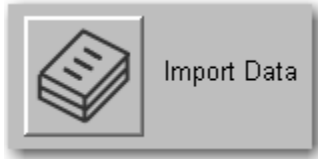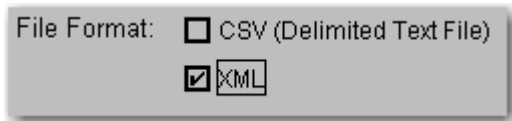
XML Import template from Studio for your collection

Next you will need and XML import template to be used as a guide for your data import.
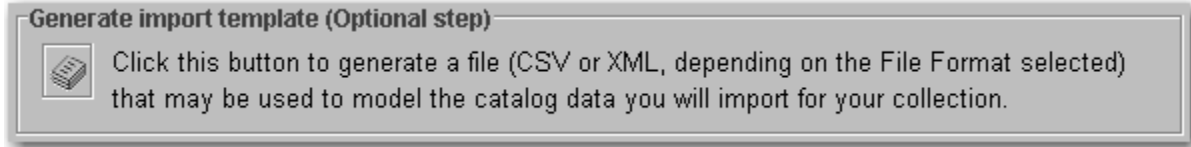
- Launch Insight Studio and login to the collection you want to import data to.
- Select the Import Data task.

- Select XML from the File Format section.

- Select the Generate import template button and save your XML import template.

The XML Import template will be similar in format to the one below, but it will be based on your Catalog Template that is used for the collection you are logged into.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!--
  - XML File generated by Insight Studio
  - Copyright (C) 2009 Luna Imaging
-->
<recordList xmlns="http://www.lunaimaging.com/xsd">
  <record type="Simple Labels" >
    <field type="Identifier" ></field>
    <field type="Creator" ></field>
    <field type="Date" ></field>
    <field type="Title" ></field>
    <field type="Description" ></field>
  </record>
</recordList>
```

An example of a populated XML Import template is located in the example folder where you installed the Insight Command Line Importer.

This template is based on the Simple Labels template. If you build a test collection with the provided Simple Labels template, this example can be used to test your system.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!--
   - XML File generated by Insight Studio
   - Copyright (C) 2008 Luna Imaging
-->
<recordList xmlns="http://www.lunaimaging.com/xsd">
  <record type="Simple Labels" >
    <field type="Identifier"
>S:\Example of dataimporter\example\Length rivers of Scotland.jpg</field>
    <field type="Creator" >Thomson, John</field>
    <field type="Creator" >Lizars, William Home</field>
    <field type="Date" >1832</field>
    <field type="Title" >Length, rivers of Scotland.</field>
    <field type="Description" >A comparative view of the lengths of the
principal rivers of Scotland.  (with) Comparative view of the height of the
falls of Foyers and Corba Linn.  Drawn by W.H. Lizars.  Engd. by W.H.
Lizars.  Published by John Thomson, Edinburgh, 1831.  (1832)</field>
        <media>
        <localMediaElement
src="file:///S:\Example of dataimporter\example\Length rivers of Scotland.j
pg" media link key="Identifier" />
    </media>
  </record>
</recordList>
```

You will notice there is an additional Media section in this document.

```xml
<media>
        <localMediaElement
src="file:///S:\Example of dataimporter\example\Length rivers of Scotland.j
pg" media link key="Identifier" />
    </media>
```

If you do not include this section, only the Data Records will be imported.

Including Media in your Import

If you will be importing Data and Media, you must follow these strict guidelines or the import will fail.

In the <media> block the src must be a full URI to the media item. So the file:/// portion of the src is critical and depending on your OS platform the path must be exact with the appropriate slashes.

Also notice the media_link_key. This must be a field in your template that will contain the same path information. In this example we used "Identifier" but you can use any text field in your template.

Notice that this field value is the exact same as the src minus the file:/// prefix or on a Unix like system file:// where only two slashes are removed.

Running the Insight Command Line Importer

Once your XML Import documents are prepared, you can run the importer by executing the command as follows:

Processing a directory of documents

DataImporter –dir <path to directory containing the XML Import files>

DataImporter –dir example

Processing a specific file

DataImporter –file <path to XML file you would like to import>

DataImporter –file example/test.xml

Important note related to limitations of this importer

At this time you can only import new records. There are no functions of this importer to keep you from duplicating existing data records. If you run this importer twice with the same source data, your result will be 2 records in your Insight Collection with the same information. You may also duplicate the Media and/or links between the records and media.

There are also no current automated ways of deleting or updating existing records using this utility.

Appendix

Insight's Data Schema

Insight's XML Data Schema is quite simple, and has a few basic elements. These elements include:

Element Name	Description
`<recordList>`	The recordList element is an optional element which can contain one or many records. Each recordList element is imported in a single media batch and import transaction. If one record fails within a list, it will not stop importing the rest of the list. *If you are only importing one record at a time, you do not need to use this element.*
`<record type="record type">`	The record element contains the basic components of a record including fields, field groups, and media elements. **type**: The Type attribute is the name of the cataloging template that will be used in the Insight Collection.
`<fieldGroup type="creator">`	Field groups represent logical groups of fields, or entities, within Insight. Field groups can contain both fields and other field groups. **type**: The Type attribute is the name of the field group or entity in the corresponding cataloging template.
`<field type="creator name">`	Field elements are the lowest level data elements. **type**: The Type attribute is the name of the field in the corresponding cataloging template.
`<media>`	The media element is a wrapper that contains localMediaElements. One will be needed for each image, audio file, and video.
`<localMediaElement src="file:///Users/luna /Desktop/Last04.jpg" media_link_key="ImageUID" />`	Each local media element represents one media file to be imported. localMediaElements contain the following attributes: **media_link_key**: the name of a unique identifier in the record. Commonly this is a field whose value is the filename and/or path to the image. This is used to link the image and data record during import. **src**: the src element is used to provide the path to the file to be imported.