Your guide to the **Content ETL Platform**

The toolbox for every content professional

xillio

making content work

The Content ETL Platform
Discover the basics

A content integration solution created for dealing with unstructured content. In its most basic form, It leverages the power of a specially designed scripting language, called Xill, to connect to systems, and it also provides a variety of administrative, analytical and enhancement tools to work with the data that is processed. As a platform, it offers a standardized solution and proven implementation methodology that enables SI's and other professionals to perform analysis, migration and integration projects in a controlled and manageable way.

Xill – The Foundation

Our in-house developed scripting language, Xill, is the basis of the platform. Xill has been designed from the ground up to be lightweight, easy to learn, and easy to maintain.

It allows anyone from young professionals to senior developers to rapidly implement connectors, and create business logic as part of their implementations.

The language comes with dozens of built-in functions which make it easy to work with all sorts of data types, and connect to a large variety of systems.

The real advantage is that you are able to skip most of the boilerplate code that usually comes with these type of implementations.

Let's take a quick look on the next page how that boilerplate code can be reduced with an example of connecting to a database. In a traditional programming language, you need to take care of a quite a few things such as SQL-flavor specifics, exception handling and the closing of result sets once you've completed working with them. We'll compare some Java code against the Xill code to perform the same functions.

Database connectivity is just one example. Xill comes with over 200 functions that let you connect to external sources and handle content in a smart and convenient way. The focus of your development effort shifts from dealing with low-level APIs to actually implementing the business rules required for your project.

Java Code

```java
package nl.xillio.contentplatform;

import java.sql.*;

public class DBSample {
    static final String JDBC_DRIVER = "com.mysql.jdbc.Driver";
    static final String DB_URL = "jdbc:mysql://localhost/EMP";
    static final String USER = "username";
    static final String PASS = "password";

    public static void main(String[] args) {
        Connection conn = null;
        Statement stmt = null;

        try {
            Class.forName("com.mysql.jdbc.Driver");
            conn = DriverManager.getConnection(DB_URL, USER, PASS);
            stmt = conn.createStatement();
            String sql = "SELECT id, first, last, age FROM Employees";
            ResultSet rs = stmt.executeQuery(sql);
            while (rs.next()) {

                /*
                 * // Here we do things with our actual query results
                 */

            }
            rs.close();
            stmt.close();
            conn.close();
        } catch (SQLException se) {
            se.printStackTrace();
        } catch (Exception e) {
            e.printStackTrace();
        } finally {
            try {
                if (stmt != null)
                    stmt.close();
            } catch (SQLException se2) {
            }
            try {
                if (conn != null)
                    conn.close();
            } catch (SQLException se) {
                se.printStackTrace();
            }
        }
    }
}
```

Xill Script

```
database("localhost", 3306, "EMP", username, password);
foreach(_, row, query("SELECT id, first, last, age FROM Employees")) {
  /*
  // Here we do things with our actual query results
  */
}
```

Connectors
Share and Receive

The WCM/ECM landscape is littered with hundreds of systems. A connector you build for your current project will most likely be valuable for someone else as well. Imagine starting your project, and already having the connectors available for the systems you're going to work on!

A sharing community

Xillio actively encourages its community to share connectors and work together on improving them. Having a large pool of ready-to-use connectors helps speed up the implementation phase greatly, as you will not have to figure out the APIs of your source or target systems – your work can be focused on making the small configuration changes to adapt the connector to the customer's specific environment.

Connectors sampling

A sampling of available connectors is included below. Please note that this list is not comprehensive, as it is being updated constantly!

SDL

sitecore

elastic

OPENTEXT

FileNet

CQ5

nuxeo

File Shares

IBM Lotus

Connect

SharePoint

ORACLE

SERENA

Livelink

Drupal™

EMC² documentum

Alfresco

HIPPO

umbraco

aDAM.

One Data Model
The Glue

Xill integrates closely with the rest of the platform by leveraging one data model that can be used for all systems. This standard data model enables us, together with our community, to share ready-to-use connectors and plugins to fetch, analyze, and enhance unstructured content. The standardized data model allows you to build modular applications with prebuilt building blocks, so that the focus of the implementation can shift to configuration and implementing the project specific requirements of the client.

In Xill script itself, adding a document to the database is as simple as creating the following code:

```
storedocument("your_content_type", {
  title = "",
  author = "",
});
```

Documents stored via the API are automatically validated, ensuring data integrity throughout the entire process of migrating content from a source to target. Once content is stored in our database, the data can be inspected, and can be enhanced or modified by modules running in the Xill Server environment. As a final benefit, all data is versioned so that a rollback can take place at any moment in case there are issues or the logic needs to be modified.

Modular
Build and Extend

The entire Content ETL Platform is built in a modular, extendable way. We include an open API which can be used to extend the platform with your own functionality or interfaces, and it can be completely customized to fit your specific requirements.

The Xillio marketplace provides our customers and partner community with a number of commercial plugins which have already been developed. These offer high-grade functionality such as auto classification and version recognition, and many more are currently under development. Finally, we provide the flexibility to develop your own plugins as well.

In addition to the Java API, the platform also provides a fully featured REST API. This means you, or our implementation partners, can build a high quality package based on the Content ETL Platform, and this can seamlessly integrate into the larger solution as needed.

Exploration
Get to Know Your Data

Knowing your data means understanding your data. The web interface of the platform contains a content browser. This browser contains multiple views through which you can gain a better understanding of the characteristics of your content, as well as the ability to edit the content directly.

This exploration feature gives you valuable insight (such as figuring out what content is viable for migration) and helps detect potential issues (e.g. by testing migration rules) early in the project, reducing overall implementation timelines and limiting errors. The platform ships with a set of pre-installed views and actions, but again, this set can be expanded through the plug-in system as necessary.

Mission control
Schedule, Run and Monitor

In a production environment you need a solid, bullet proof environment in which to run your processes, particularly because in many IT environments you may not directly have access to the production system. The Content ETL Platform comes with standard, built-in functionality to run, monitor and interact with Xill scripts and modules.

Scripts can be scheduled to run on a regular basis or at specific intervals, they can be set up to run with dependencies and in a certain order, or even multi-threaded if required. All server and script activity is logged to a database, and can be queried through the web interface or through your own business intelligence application for easy reporting and monitoring of the processes.
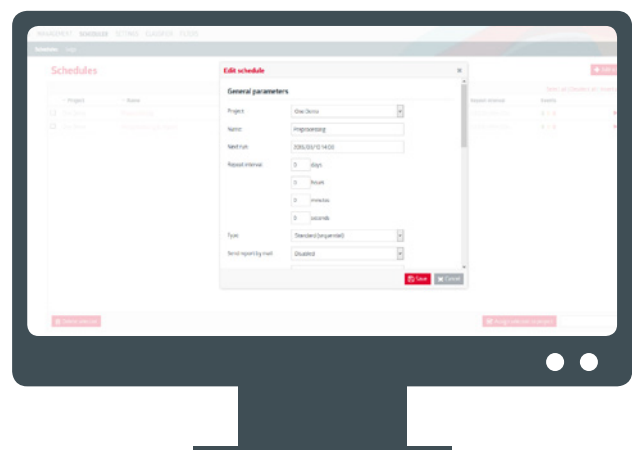
IDE
Deploy & Go

The entire content platform is built to be easy to use, and easy to deploy. We have built it this way intentionally, because in a project your primary concern is to help the customer and not to "waste" time setting things up. The Xill scripting language is no exception to this rule. It comes with a fully featured IDE including a debugger, extensive help, code examples, and close integration with the rest of the platform.

Our IDE has a one-click installer, requires no configuration and has all the tools you need to do your job in the right way, quickly, and in a controlled environment. You can literally begin working within one minute! For senior developers looking to plug in Xill into their existing IDE, keep an eye on our roadmap for future developments!
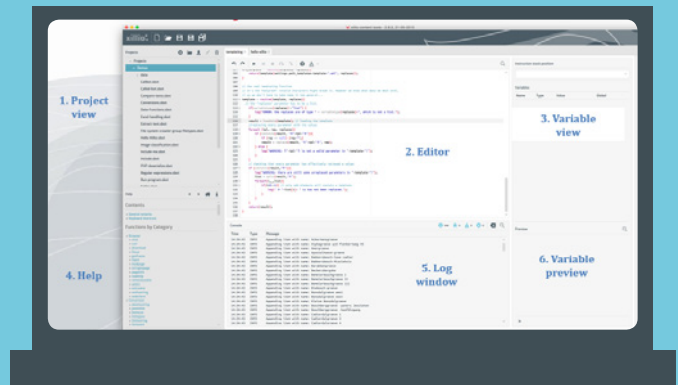
Technology
Our Stack

The primary criteria for our choice of technology stack is that we want our tools to be OS independent. After all, our client's systems run on all different flavors of operating systems, and we need to be able to handle those differences. Because of this, we chose Java as our main platform language of choice. For those interested, we use of technologies such as Java, Spring, Hibernate, AngularJS, Elastic Search, MongoDB, MySQL, Selenium, Python, Tika, POI, Saxon and many others.

System requirements
Up and Running

The system requirements for the platform are very reasonable. We advise the following (minimum) system specifications:

- OS: 64 bit Windows, OSX, or Linux
- CPU: Intel i7 quad core or equivalent
- Memory: 8GB
- Disk space: 1GB for software, plus as much GB as you need for local storage of data (50GB minimum suggested)

Module Preview
Classification

One of our most prominent modules, the Xillio Classifier provides a whole new way of enhancing the metadata of your document set. The classifier hosts a state-of-the-art piece of artificial intelligence (AI) that can be trained to recognize document characteristics such as topic, keywords, document type, or any other taxonomy that may exist in your ECM system.

The classifier comes with built-in functionality to not only train the AI engine, but also allows you to do test runs of the results via a friendly, web-based user interface. The classifier supports over 25 languages out of the box, meaning it can be deployed in multilingual environments just as easily. As a final benefit, there is full integration with the Xill script language so that any business rules defined in Xill can be added to the classifier to further enhance the quality of the result set.
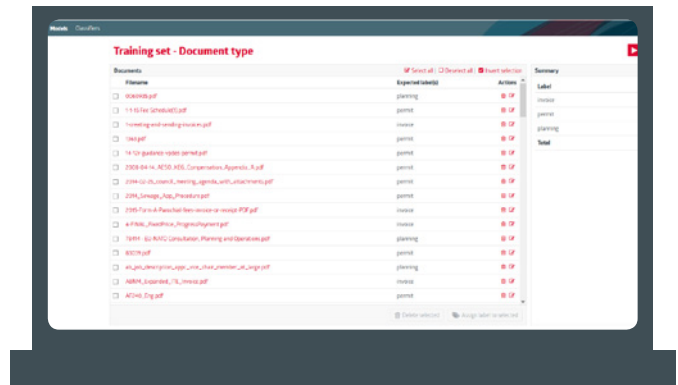
Deduplication

On average, 10 to 20 percent of the documents in collections we see, are duplicates, and in some cases it is substantially more, sometimes upwards of 50 percent.

Our deduplication module is another cool piece of engineering as it not only finds identical duplicates (which is what you would expect it to do), but can also look at the content within a document to determine how closely alike they truly are. By obtaining this information, you can align all of the versions of documents, and remove those which are polluting your system. Our algorithms scale linearly, and will work on normal consumer hardware for large numbers of documents, even in the millions. The integration with the Xill script language allows you to define more granular rules, which will help determine which documents to process and which to remove.
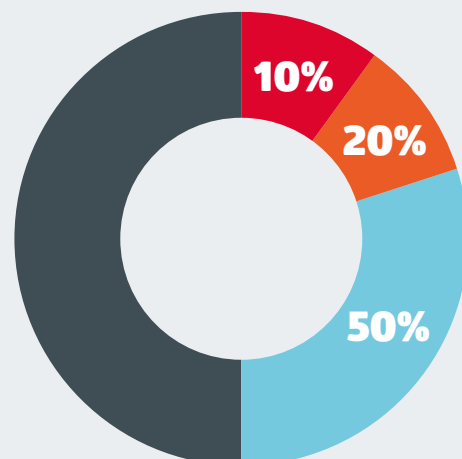
Structure Mapping

Our Structure Mapping module lets business users restructure their documents and folders in a completely sandboxed environment. This allows our customer to be actively involved during the migration, and to influence and create the required structure at the same time.

With this module the user is presented a "structure proposal" in a sandboxed environment, which they can fine tune without affecting the live production environment. During the migration process, documents are then loaded into the target system using the newly defined structure. The process becomes safer, and requires significantly less work by the business users because there is little to no rework required.
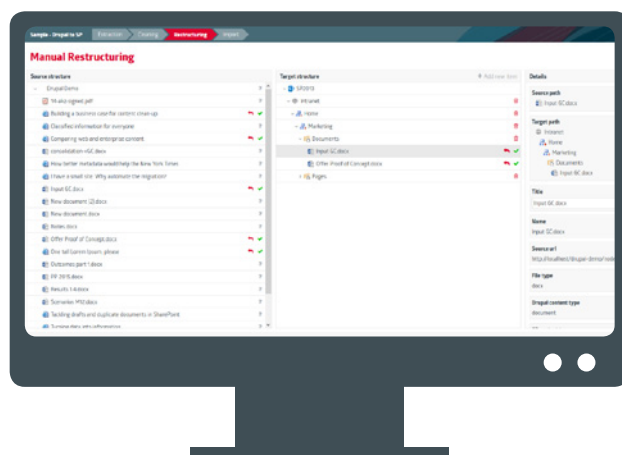
Content Analysis

The Content Analysis module lets you take a deep dive into the content of documents. It lets you search for specific patterns, so that you can more easily detect those documents that need attention. For example, you could use it to find all documents containing a credit card number, or other sensitive personal identifying information. Based on this analysis, you can use the rest of the Content ETL Platform to apply the appropriate business logic to deal with these documents as necessary.

Contact us

Netherlands . Julianalaan 15 . 1213 AP Hilversum
the Netherlands . +31 35 622 95 45

USA . 300 North LaSalle . Suite 4925
Chicago, IL 60654 . +1 (312) 498-1816

info@xillio.com . www.xillio.com