IRI NextForm
Data, File & Database Migration

Product Overview

Technical Summary, Samples, and Specifications

IRI | Total Data Management | CELEBRATING 40 YEARS 2018 | The CoSort Company

Table of Contents

NextForm Introduction

[IRI NextForm](#) is a data reformatting, replication and reporting package for data stored in legacy and modern databases, index and sequential files, and dark data (unstructured) documents. NextForm does not need Hadoop, in-memory DBs, or more hardware to process big data, but it can migrate data to and from legacy and modern platforms. There are no limits on the number or size of sources or targets.

NextForm is the only data migration tool that combines the power and simplicity of CoSort's high-volume data mapping engine with the familiar ergonomics and extensibility of Eclipse.

NextForm is a multi-edition product in the _IRI Data Manager_ suite designed to rapidly:

- Remap data types, record layouts, file formats, and endianness
- Convert column data, layouts, and relationships (constraints) between databases
- Replicate data from one or more sources to similar or different table or file targets
- Federate, or virtualize, data by producing ad hoc views from disparate sources
- Report on data from one or more sources, or create filtered hand-offs for BI tools

NextForm Attributes

Speed - NextForm uses the CoSort engine for big data movement and manipulation. This means that you can convert massive file and table data sources -- plus reformat and report on data in highly customized ways -- all at the same time. There is no code compilation, nor need for multiple job steps, I/O passes, or hardware partitions.

Versatility - NextForm handles table, file, record, and layout changes. It supports all popular databases, more than a dozen sequential file formats, unstructured textual sources, and more than one hundred data types.

Ease of Use - NextForm is supported in an Eclipse GUI called IRI Workbench. Its simple metadata discovery and format conversion wizards quickly step you through data and re-format definitions. The GUI creates text scripts that you can easily modify and re-use.

Interoperability - NextForm uses the same 4GL metadata and Eclipse GUI to define, effect, source, and target mappings as IRI CoSort (for data transformation), IRI FieldShield (for data masking), and IRI RowGen (for test data generation). The NextForm data definition file (.ddf) format is also supported by the Meta Integration Model Bridge (MIMB) and AnalytixDS Mapping Manager technologies so you can more rapidly create migration jobs using metadata already defined for the third-party data manipulation, modeling, and reporting tools those hubs support.

Features	Benefits
Reformats huge files, reliably migrates legacy data sets to client/server RDBs	Facilitates right-sizing, reduces data synchronization needs, and saves time and money over custom coding (plus acquisition, implementation, and support of) legacy platforms
Converts over 100 data types	Allows data re-use in new platforms and applications
Structures unstructured data	Extracts data values and shows associations for relational use
Re-maps field layouts	Supports pivot transformations, reporting needs, and database repopulation (inserts)
Data replication	Allows re-use of same or re-formatting data copies without new system re-generation, and facilitates data virtualization
Migrates database versions and platforms	Frees data and application from stale procedures,and frees company's locked into specific vendor terms
Built-in report generation	Migrate and expose information in multiple sources in custom displays suitable for billing and analytics
Interoperable metadata	Re-use source and target data definitions, as well as mapping and manipulation specifications in IRI CoSort (for data transformation and reporting), IRI FieldShield (data masking), and IRI RowGen (test data generation)
Simple data and definition (metadata) syntax	Data migration tasks are easy to configure, run, modify, and share
Cross-platform; runs on Windows, Unix/Linux	The same jobs defined or run on one platform can be modified or run on any other
Low cost licenses and volume discounts	Immediate project ROI and low TCO for enterprise-wide deployment

NextForm GUI

The IRI Workbench GUI for NextForm supports the discovery and definition of disparate data sources in both local and remote systems. Its profiling, pattern searching, forensic metadata discovery, and metadata management tools support NextForm and other IRI software (upgrade) features, including: data integration, data masking, data quality, and data curation.

Use the GUI to create, edit, and manage NextForm control language (.NCL) jobs. NCL job scripts are simple, self-documenting text files that define the layout of your data sources and targets. Save jobs within projects, and run them from the GUI. Or, run them from the command line, batch program, job scheduler, or your application.

The GUI also supports the specification and re-use of data manipulation rules you can apply to multiple tables with pattern-matching column names.For example, standardize a date calculation, unify column data types, specify a custom mapping, or insert a lookup/master value.

Data Discovery

To migrate, federate, or report on data, basic awareness of its structure and content is essential. The included data discovery tools in the IRI Workbench GUI for NextForm:

- produce statistical information about data
- reveal and validate relationships between entities
- find and extract data that matches patterns, fuzzy search, or dictionary (set file) values
- expose forensic metadata
- discover or auto-define source metadata

across disparate data sources, including database tables, flat files, and dark data (unstructured text) documents.
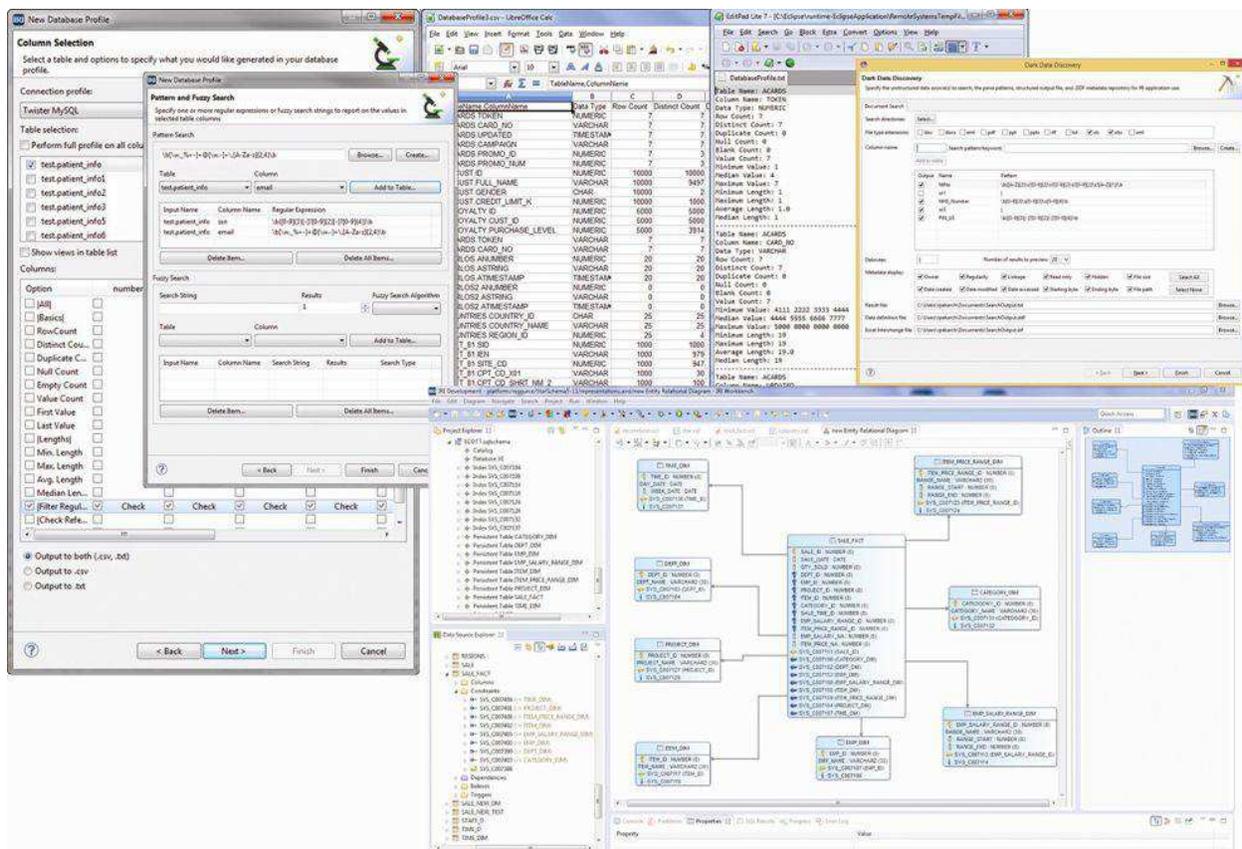
Data discovery processes also support data cleansing, reformatting, and reporting. These tools can help also you examine the structure and completeness of database data, and validate that the proper data is being stored in the right places.

Data and File Format Conversion

IRI NextForm can help you convert the essential format, record layout, data types, and endian characteristics in one structured data source (file or table) to another. This is helpful not only for quick, one-off file migrations, but also for data replication, and possibly federation.

The GUI's "*New Single Source Migration*" job wizard shown below converts the format of one data source (e.g., an LDIF file) to a target file in the same or different format (e.g., CSV):


This result of this process is a portable job script in the NextForm Control Language (NCL) which you can modify in the GUI or any text editor. From this job specification, you can also remap field positions and sizes, change data types, endian characteristics, and so on -- either directly in your .NCL script or the output field layout editor, a GUI dialog.

Database Migration

NextForm users migrating between database versions or vendors should run the "*NextForm Multi-Table Migration*" job wizard in the IRI Workbench GUI for NextForm.

Use it to connect to the source tables and define mapping rules to apply across related columns. Create and specify the structure and relationships of your new tables automatically.
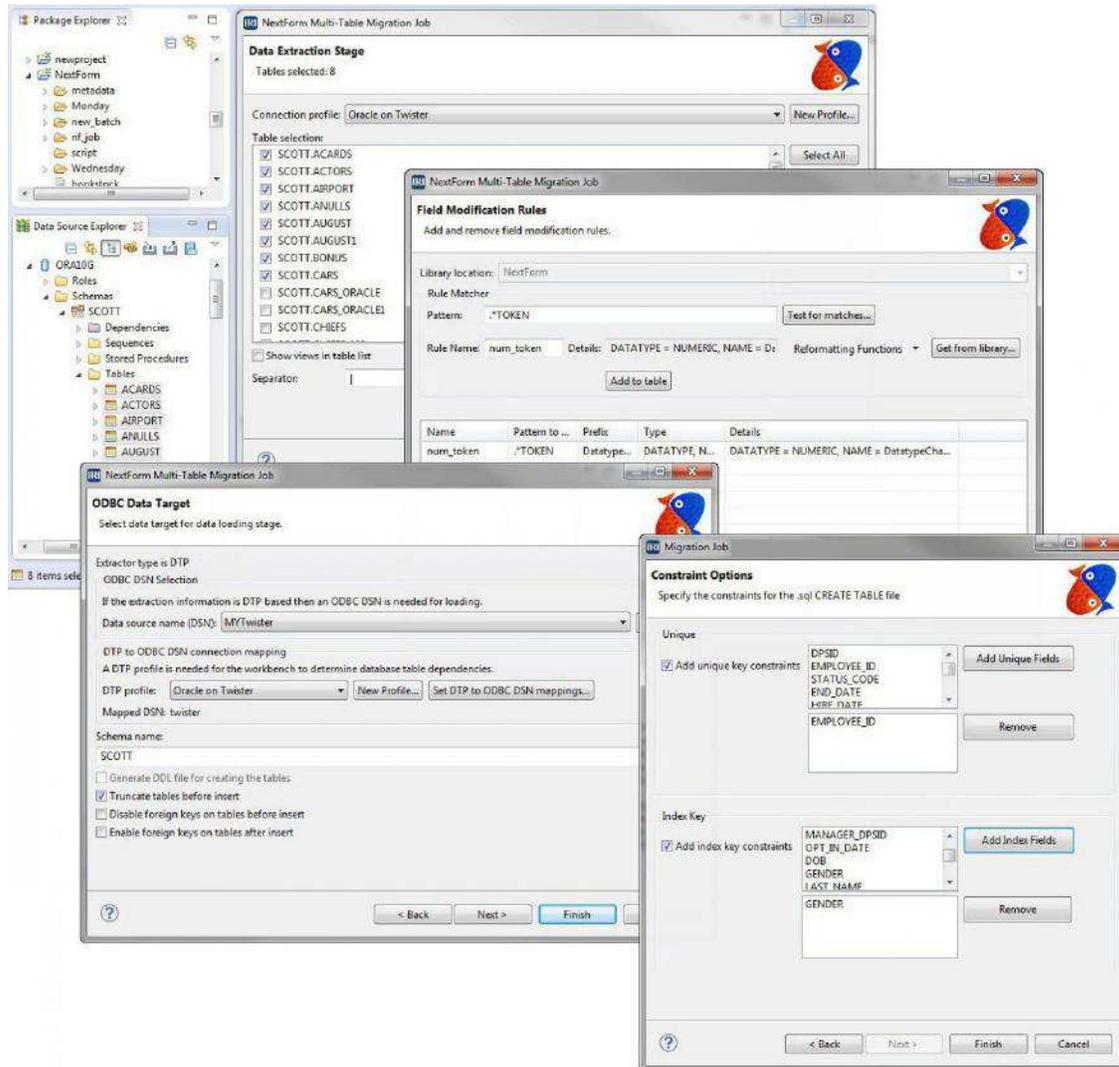
Populate the tables with ODBC in the wizard, or on your own via JDBC or DB load utilities, also supported in the GUI. All .NCL job scripts for data acquisition, manipulation, and mapping to targets are produced automatically. Modify them by hand in the GUI or in any text editor.

Use IRI CoSort to easily add data transformation, masking, and reporting functions into the process. Remember, NextForm runs in the same GUI, and its job scripts use the same metadata as other IRI tools, so enhancing these operations is seamless.

Structuring Unstructured Data

All NextForm edition users can run the *Dark Data Discovery Wizard* in the IRI Workbench GUI to search, parse, forensically examine, and structure "dark data" in unstructured text files.

The wizard has you specify string literals, patterns (via expression logic), or a data dictionary (set file) containing the values you want to find. The search allows you to preview the data it discovered in the file formats you selected on your local computer or network, and then saves it all in a flat file along with an IRI-compatible data definition file (DDF) metadata repository.


Having the discovered values in a flat-file structure can help reveal new meaning and associations and facilitate the use of NextForm for mapping, replicating, and reporting on the data now in flat files. The results of the search can also reveal additional forensic metadata about the file sources, including ownership, creation and modification dates, sizes, links, etc.

Data Replication

You can also use NextForm and its Eclipse GUI for [replicating](#), or copying, data (or filtered subsets of data) in the same or different formats.

Connect to disparate data sources, select the contents you need, and copy those field values to one or more targets, in the same or different formats.
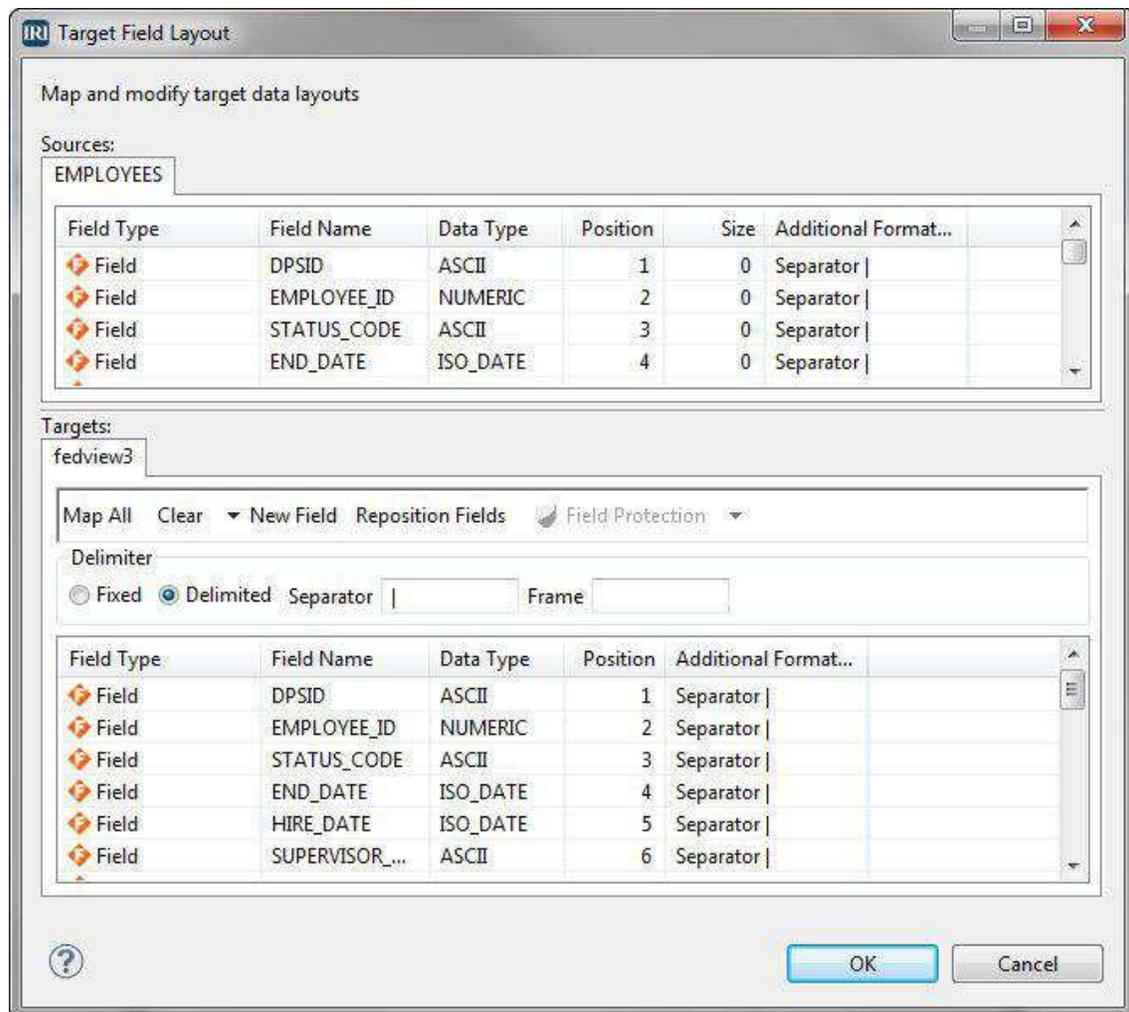
This dialog is used within NextForm data migration job wizards, or opened independently from any existing job script in the Workbench editing window. You can also specify new data sources, targets, and field-level manipulations in the dialogs, or directly in the job scripts.

Upgrade to an [IRI Voracity](#) subscription for more complex, ETL-style data replication, as well as data cleansing, data masking, database subsetting, and master data management. Voracity uses the same metadata and GUI as NextForm so no changes to existing jobs are necessary.
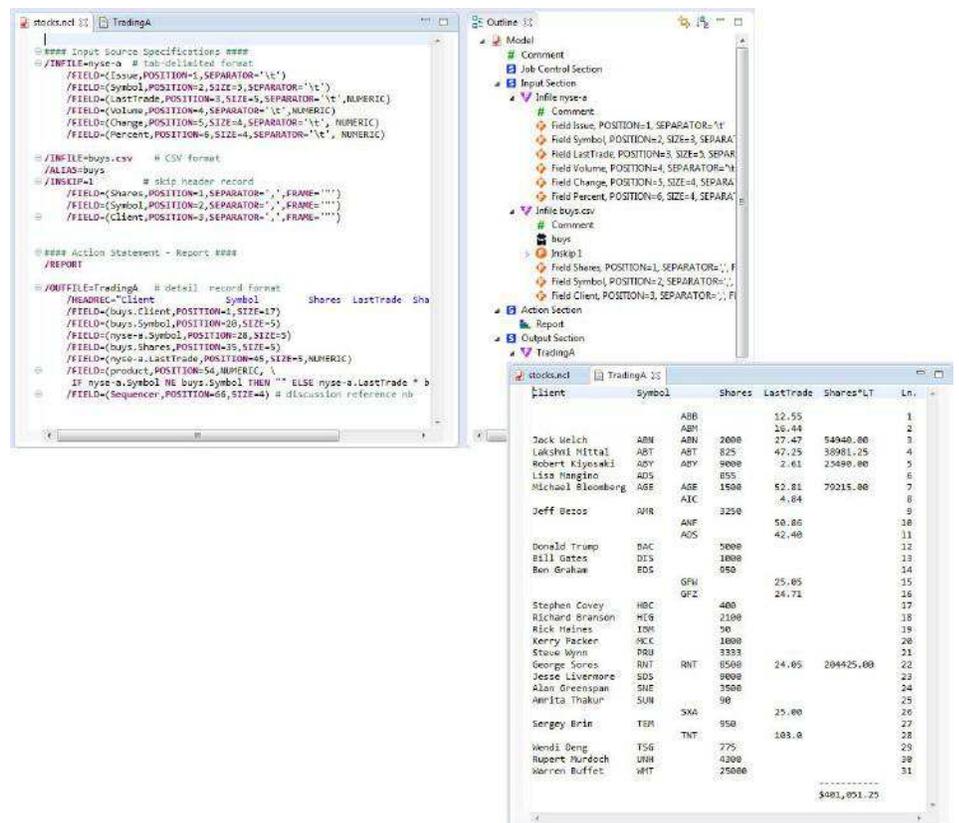
Data Federation
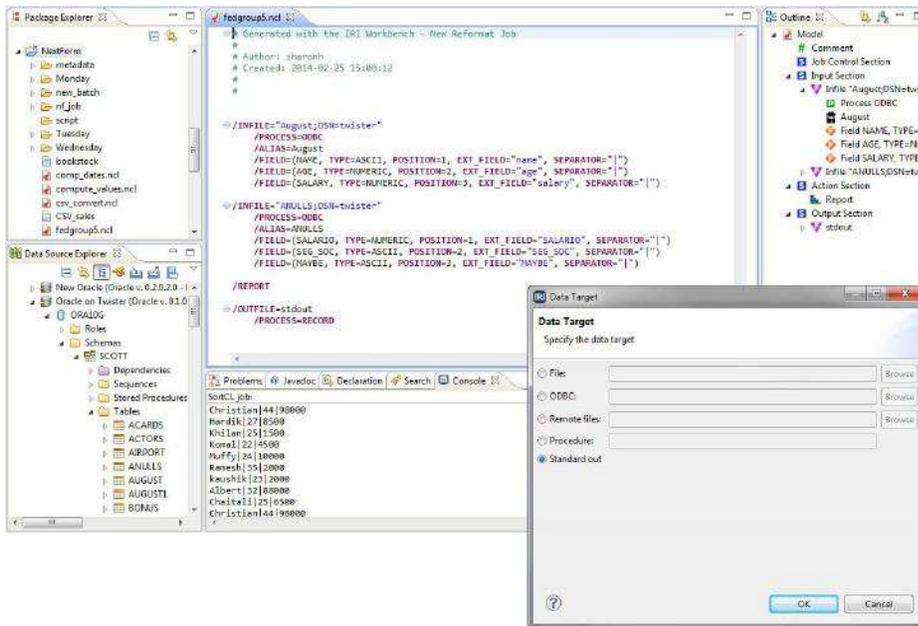
Use the NextForm GUI to create virtual sets, or federated views, of data you do not need to store.

Connect to and auto-define the metadata of your data sources to blend and produce ad hoc database or file targets, including immediate console views (using "stdout") in the GUI.

Reporting

You can use NextForm and its GUI to generate meaningful, customized business intelligence reports from data in one or more sources. NextForm offers the same report generation features as CoSort SortCL.

Other reporting options include using BIRT in Eclipse to render data that NextForm produces in graphical dashboard widgets, and/or creating CSV, XML, and ODBC targets for third-party BI and analytic tools.

NextForm Editions

Lite (Free) *Featuring NextForm executable and IRI Workbench (Eclipse GUI)* Migrate, replicate, federate, and report on flat files. Convert fixed to delimited, LDIF to CSV, EBCDIC to ASCII, big to little endian, etc.	**COBOL** *Featuring Lite + MF Libraries* Convert MF-ISAM & Vision Files, MF & RM data types. Report from and re-use mainframe data sets.
DBMS *Featuring Lite + ODBC Libraries* Replicate, federate, and report against data in any connected table or flat file. Upgrade or migrate from/to Oracle, DB2, Sybase, MySQL, SQL Server, Teradata, and other DBs.	**Legacy** *Featuring Lite + Proprietary Drivers* Access and view legacy data in over 120 different formats.
Modern *Featuring Lite + Proprietary Drivers* Access and migrate data in big data and cloud and SaaS repositories.	**Premium** *Featuring Lite + All Libraries* Identify, integrate, use, and report on structured, semi-structured, and unstructured data.

NextForm-Compatible Applications

NextForm uses the same metadata as other IRI software. The source and target data definitions established for use in NextForm can be used in any other IRI software product. Third-party metadata mapping tools also support NextForm's data layout and manipulation syntax.

- IRI Voracity - Data Discovery, Integration, Migration, Governance and Analytics
- IRI CoSort - Data Sorting, Transformation and Reporting
- IRI FieldShield - Data Masking & Encryption
- IRI RowGen - Safe, Intelligent Test Data
- MITI MIMB - Meta Integration Model Bridge
- AnalytiX DS - Mapping Manager

Technical Specifications

IRI NextForm is a general-purpose data and database conversion tool with built-in data discovery, reformatting, replication, and reporting capabilities. NextForm uses the same data definition and mapping metadata as the IRI CoSort 'Sort Control Language (SortCL)' program, and the same Eclipse GUI for all IRI software, IRI Workbench. NextForm includes the features listed in this matrix, and thus many other features exposed through the Workbench GU for all IRI software. Following is a non-exhaustive list of the functionality supported in NextForm:

Installation

- Distributed via internet or user-specified media
- Menu-driven setup and configuration utility for the CLI

Invocation

- Command line (including pipe sequences), shell commands, and batch scripts
- IRI Workbench (Eclipse GUI) 'Run As' menu, 'Run Configuration', or task scheduler
- Third-party automation (scheduling tools) like Stonebranch Universal Controller

Data Discovery (Profiling)

- Searches for user-specified literals (string values), strings in a dictionary look-up file, or values conforming to patterns expressed in regular expression syntax across flat-files, JDBC-connected data sources, and "dark data" document file formats
- Extracts values and metadata from document file formats (HTML, MS Office, PDF, RTF, TXT, XML) based on regular expressions into a flat file and creates an IRI DDF file for it
- Generates statistical reports on selected attributes of flat-file and JDBC data sources
- Validates relationships (checks for referential integrity) between JDBC-connected tables
- Produces E-R diagrams for relational databases connected through JDBC
- Auto-creates, saves, and uses field definitions in IRI DDF syntax for file and DB sources

Input and Output

- Processes any number of files and ODBC sources listed here, of any size, and any number of records, fixed or variable length (to 65,535 bytes) passed from those sources, an input procedure, from stdin, a named pipe, a table in memory, or from an application
- Supports environment variables and wildcards in source and target specifications, along with absolute and relative path names, ODBC DSN files, aliases, and URLs (pending)
- Accepts and outputs fixed- or variable-length records with delimited fields
- Creates detail and unsorted summary reports, plus hand-off files or tables for BI tools
- Outputs sequence (index) numbers in each row, at user-defined start and interval values
- Builds set files from DB columns or user-defined, composite formats and value ranges
- Returns processed (transformed, masked, formatted, etc.) records one (or more) at a time to a file, pipe, output procedure, formatted report, database table, and/or application

Data Sources

Via default or partner drivers, NextForm migrates and processes these sources and targets:

Databases
• Oracle • DB2 UDB • DB2 Mainframe² • MySQL • SQL Server • Sybase • Teradata

Data Type Groups
• ASCII (C types) • Numerics (All) • EBCDIC MF & RM COBOL • Endian (Big, Little, BOM) • Dates • Timestamps • Composite • Single-Byte • Multi-Byte (CJK, Thai) • Unicode (UTF-16) • Packed Decimal • Zoned Decimal

File Formats
• CSV • Delimited Text • Fixed-position Text • LDIF • Micro Focus Variable Length • Micro Focus ISAM¹ • Line / Record / Variable Sequential • Variable Blocked • Vision¹ • VSAM, QSAM, KSAM, C-ISAM² • XLS (Excel) • XML³

Legacy / Mainframe
• Access • Adabas • Advanced Pick • ALLBASE • Alpha Five • Amazon RDS • Azure • BizTalk • Clipper • Codasyl • CorVision • ConceptBase • D-ISAM • D3 • Datacom • Dataflex • Db4o • dBase • Desktop Adapter • DL/1 • DSM • Enscribe • Enterprise Adapter • FileMaker • Firebird • Focus • FoxPro • GA-Power95, R91 • Gemstone • GENESIS • Gigabase • H2 • IDMS • IDS • Image • IMS • InterBase • Intersystems Cache • ISM • Jasmine • jBase • K-ISAM • Knowledgeman • KSDS • Lotus Approach/Notes • Manman • Mentor • Model 204 • Mumps • MyBase • Netezza • NonStop • SQL • ObjectStore • Paradox • Pathway • PDS • PervasiveSQL • Pick/Pick64+ • PI-Open • Powerflex • Powerhouse • Progress • QueryObject • rBase • R83 • Rdb • REALITY • Red Brick • RMS • Reality/X • RRDS • SAP HANA • Sequoia • Sharebase • Supra • Terracotta • Total • Ultimate • UltPlus • UniData • UniVerse • VSAM

Modern / Big Data
• Amazon EMR Hive • Apache Cassandra • Apache Hadoop Hive • Cloudera CDH Hive • Cloudera Impala • Database.com • FinancialForce • Force.com apps • Hortonworks Hive • Hubspot • Lightning Connect • MapR Hive • Marketo • MongoDB • MS Dynamics CRM • MS SQL Azure • Oracle Eloqua • Oracle Service Cloud • Pivotal Greenplum • Pivotal HD Hive • Salesforce.com • ServiceMAX • Spark SQL • Veeva CRM

Unstructured Sources
• ASN.1 TAP3 • DOC, .DOCX • EML, .OST, .PST • PDF, RTF • PPT., .PPTX • TXT, .XML • XLS, .XLSX

Record Selection and Grouping

- Includes or omits source and target data by comparing fields to each other or constants, or through SQL /QUERY (SELECT) logic
- Skips a specified number of records, bytes, or a record header
- Processes a specified number of records or bytes, including a saved header
- Supports any number of fields from 10 to 65,535 bytes in length
- Orders fields in fixed position or floating (on one or more delimiters)
- Supports numeric data types, including all C, FORTRAN, and COBOL data types
- Supports single and multi-byte character keys, including ASCII, EBCDIC, ASCII in EBCDIC sequence, Thai characters, and natural (locale-dependent) values
- Supports American, European, ISO and Japanese timestamps
- Supports Unicode and double-byte characters like Big5, EUC-TW, UTF-32, and SJIS
- Allows left or right alignment and case shifting of character keys
- Accepts user compare procedures for multi-byte, encrypted and other special data
- Controls treatment of null fields when specifying floating (character separated) keys
- Collates (and converts between many of) the data types (formats) described here.

Record Reformatting

- Inserts, removes, resizes, and reorders fields, or columns within rows where permitted
- Derives new field values through the use of various field-level functions
- Converts data in fields from one format to another either using internal conversion
- Maps common fields from differently formatted input files into one or more targets
- Performs math operations and functions on field data to generate new output field values
- Changes record layouts from one file type to another, including: Line Sequential, Record Sequential, Variable Sequential, Blocked, Microsoft Comma Separated Values (CSV), ACUCOBOL Vision, Micro Focus I-SAM, MFVL, Unisys VBF, VSAM, W3C Extended Log Format, LDIF, and XML
- Maps processed records to many differently formatted output files
- Writes multiple record formats to the same file for complex report requirements

Data Cleansing, Reformatting, Protection and Validation (Governance)

- Creates and processes sub-strings of original field contents, where you can specify a positive or negative offset (from the left or right, respectively, of the source field) and a number of bytes to be contained in the sub-string
- Finds a user-specified text string in a given field, and replaces all occurrences of it with a different user-specified text string on output
- Analyzes fields to display the offset number for the specified occurrence of a string
- Manipulates and displays literal values with input data inside field statements for use in value derivations, functions, conditions, cross calculations, and reporting

- Aligns desired field contents to the left or right of an virtual (pre-action layout) or output field, where leading or trailing fill characters from the source are moved to the opposite side of the string
- Supports Perl Compatible Regular Expressions (PCRE), and pattern searches
- Uses C-style "iscompare" functions to validate field contents (e.g., for printability) for use in /INCLUDE and /OMIT record filtering statements
- Protects sensitive data with field-level de-identification and multiple encryption routines, as well as anonymization, pseudonymization, randomization, hashing, filtering, string manipulation, and data masking (obfuscation / character redaction) techniques
- Creates CSV, XML, or ODBC hand-offs for third-party BI and analytic tools
- Produces data for in-memory, ODA driver consumption by Eclipse BIRT at reporting time

Metadata Controls

- Generates data (field) definition files (DDF) automatically from database tables, delimited files, the metadata discovery wizard, and new job wizards
- Converts native COBOL copybook, Oracle SQL*Loader control file, CSV, and W3C extended log format (ELF) file layouts into DDF
- Modifies DDFs in syntax-aware text editor, fit-for-purpose form editor, and dialogs
- Supports the generation and modification of NextForm Control Language (.NCL) job scripts in syntax-aware text editors, form editors, and dialogs
- Supports the generation and modification of SQL select statement and procedures
- Saves all metadata assets in project folders for copying, sharing, printing, etc.
- Interacts with Git, SVN, CVS and other Eclipse-compatible version control systems.

Ease of Use

- Uses a self-documenting 4GL to define and process data
- Leverages centralized application and file layout definitions (metadata repositories)
- Incorporates all design and deployment in the familiar, graphical IDE of Eclipse
- Provides a color-coded, syntax-aware metadata (layout and script) editor
- Produces and saves optional runtime warnings
- Logs performance and application statistics, and an XML audit file, with each run
- Reports problems to standard error when invoked from a program, or to an error log
- Runs silently or with verbose messaging without user intervention
- Allows user control over the amount of informational output produced
- Supplies first steps, demo projects, and tutorial 'cheat sheets' to facilitate onboarding
- Describes commands and options through Eclipse help and .pdf documentation
- Easy-to-use interfaces and seamless third-party sort replacements preclude the need for training classes; however, advanced training is available in Florida or at user sites
- Phone, web, and email support available directly from the product developers
- Local language support is available from more than 30 international offices.

INNOVATIVE ROUTINES INTERNATIONAL (IRI), INC.

2194 Highway A1A
Melbourne, FL 32937 USA
Phone +1 321-777-8889
http://www.iri.com