

# Data Express

The reliability of IT systems is a key requirement of almost any organization. Unexpected failure of enterprise systems can be expensive and damaging to an organization. It is vital therefore that testing processes and procedures are rigorous and complete.

## Product Highlights

An essential part of the testing process is the data used during test execution. While artificially created, test data can be a valuable starting point because it can replicate the unique and unexpected characteristics of production data. Therefore, to avoid any unexpected failures when applications go live, the majority of organizations use production-derived data as part of the testing process.

However, using real data can raise challenges both in terms of data size and privacy. For both internal use and outsourcing, data must be masked, de-identified or obfuscated in order to protect the privacy of clients and ensure that organizations meet compliance requirements. Production data can also be very large. Using terabytes of data as part of the testing process can add significant delays and cost both in terms of preparing test data and executing the test process itself.

OpenText™ Data Express is a highly productive, automated and repeatable solution for the consistent creation of secure and manageable test data. Using Data Express, organizations can rapidly mask production data from across the enterprise, delivering non identifiable data for the testing process that is both compliant with privacy requirements and optimized in size.

## Key Benefits

### Delivering Regulatory and Legislative Compliance with Test Data Privacy

Data privacy legislation and industry regulations such as GDPR, PCI, HIPAA, Basel II, Sarbanes Oxley and others require that personal identity information is protected. While production systems will typically have well developed layers of security, both logical and physical to protect data from misuse, a development environment has few (if any) of these safeguards. Data Express provides a centralized automated mechanism to mask enterprise data sources. It consistently creates secure non identifiable yet realistic data that can be used during the development and testing process, delivering compliance and reducing the risk of a data breach. In addition, Data Express can provide an essential and safe training environment for end-users, when using live production data for testing or training.

Data Express is aligned with GDPR articles, strengthening and unifying data protection for individuals by:

- Detecting and identifying personal data with Data Discovery and other processes, addressing "Right to erasure" (Article 17)
- Providing "Pseudonymisation" through Data Masking and Synthetic Data Generation
- Addressing "data minimization" through Data Subsetting

### Improving Testing Productivity and Time to Market

Providing the right test data environment can significantly improve the application delivery process, and consequently the quality of business applications and overall agility of modernization initiatives. Cumbersome, manual or poorly integrated testing processes add significant time, cost and risk to the task. Testing can become a liability and not an IT asset, and can often limit IT's overall flexibility and business alignment. Data Express allows organizations to improve the testing process by rapidly and consistently providing test data when needed and in the format that it is needed. This helps developers, quality assurance teams and database administrators to work efficiently with secure, comprehensive and compliant data.

### Reduced Costs and Improved Quality

The testing process is a major element of the development lifecycle. The availability of appropriate and accurate data for system testing, unit testing and acceptance testing is critical. Manual processes to secure and manage test data can take time and resources while potentially returning inconsistent results. By providing a repeatable and automated solution, Data Express reduces the resources needed to create test data, saving time and costs in the test cycle and delivering improved quality.

Through providing a centralized approach that understands data relationships, Data Express can substantially reduce the test data size, with typical test data volumes cut by over

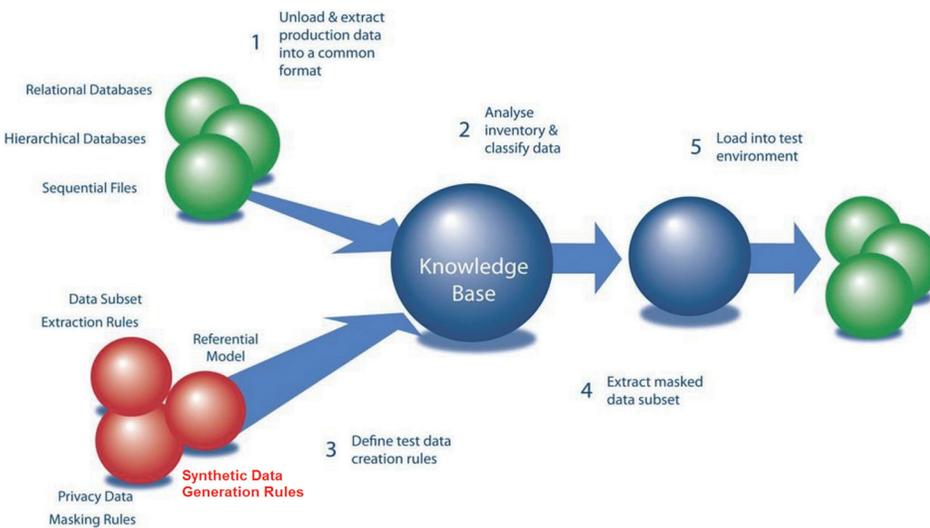


Figure 1. Strategic architecture for subsetting and masking with Data Express

90%. With such substantial size reductions, the time taken to execute the testing cycle can be reduced. Particularly on mainframe environments, significant savings in processing and storage costs can be achieved. With its consistent process Data Express can ensure the repeated reliability of test data, removing the risk of 'false positives' and enhancing the efficiency of the testing process.

## Key Features

### Knowledge Base

The builder component of Data Express allows an inventory of organizational data to be taken, collated and centrally stored in the Knowledge Base, a single meta-data repository. Strategic and business-critical application data stores, including IMS DB, DB2, VSAM, ADABAS, Oracle, SQL Server, Sybase and any ODBC accessible data store can be collated in a single place. The resulting centralized information can then be reviewed from a single administrator desktop, yielding a single view of mission critical data, a single base from which to consistently subset and mask, and a single core of information to report against. Once the

Knowledge Base is built it is not necessary to repeat this process even if masking or subset criteria change. Lifecycle support enables straightforward maintenance of the knowledge base as data stores evolve.

### Comprehensive and Customizable Masking Delivers Test Data Privacy

The data masking module supports the rapid de-identification or obfuscation of data according to a defined set of rules and criteria. After the builder component has created the central knowledge base it classifies and inventories the data enabling the definition and execution of a set of rules and actions to be taken by the data masking module to de-identify the

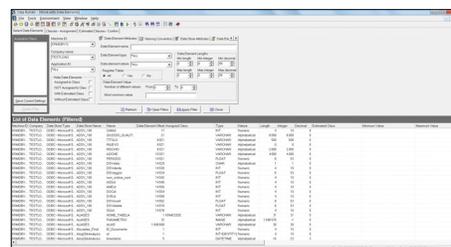


Figure 2. Building the Knowledge Base

data according to the required criteria. The data masking module facilitates the rapid generation of consistent test environments. It removes those data elements that could be used to gain additional knowledge about sensitive information and makes it impossible to trace personal information.

- For data de-identification, the builder and data masking modules allow:
- An inventory of databases and files containing sensitive data
- Rapid classification and cataloging of personal and sensitive data including fingerprinting capability. This enables selected known personal identity fields to be sampled in order to find unknown sensitive fields with the same characteristics (Data Discovery)
- The masking of personal identification data (personal details, codes, etc.) making sensitive data anonymous in test environments
- The creation of de-identified test data that respects referential integrity masking consistently across data stores for accurate and meaningful testing
- Compliance with EU, UK, US and other international data privacy laws and with regulatory standards such as GDPR, PCI, BASEL II, MIFID, Sarbanes-Oxley, and HIPAA

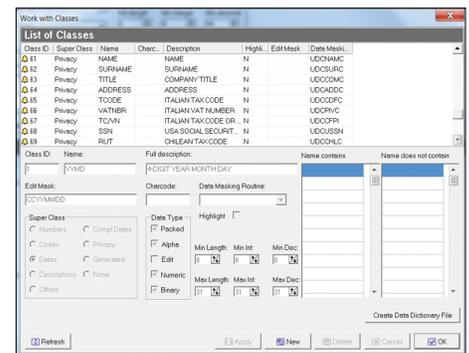


Figure 3. Class Definition with Data Express

The properties of the data masking process are:

- The use of supplied routines or customization of masking changer programs to tailor the de-identification process for individual requirements and incorporate existing masking processes if required
- Predefined masking routines which can convert name; surname; address; dates; telephone numbers; regional formats such as US SSN numbers; Italian, Spanish, and Chilean tax codes; email addresses; credit card numbers; bank account numbers; company names IP numbers; and other unique ID codes to meaningful values
- The application of same masking scheme to different fields in respect of key relationships
- Common portable masking routines across multiple platforms

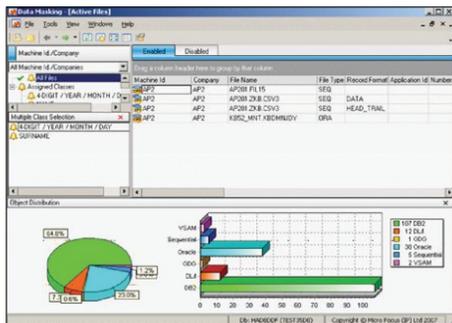


Figure 4. Data Masking Multiple Class Selection

### Rapidly Define and Create Data Subsets Delivering Faster and More Agile Testing

The data subset extraction module allows organizations to extract a consistent and congruent subset of data, to reduce the volumes of data in the test environment. This process accelerates the execution of quality controls following application changes by creating an accurate and representative test sample of the application environment. The methodology

covers environmental analysis, parameter definitions, file and layout cataloguing, data dictionary construction and master database definition. The client tool extracts the information from the repository and defines the logical model of data and rules for the sub-setting. Afterwards, the core subsetting engine applies the rules which allow a revised set of data to be written to a new database.

The properties of the data subset extraction process are:

- Multiple extraction rule definitions supporting the creation of different reduced data sets for specific purposes such as system testing, performance testing and unit testing
- Repeatable extraction schemes providing consistent results
- Extraction rules wizard to generate definitions for complex extraction patterns allowing the subsetting of data sets according to multiple criteria
- Automated extraction of related records using the same key values

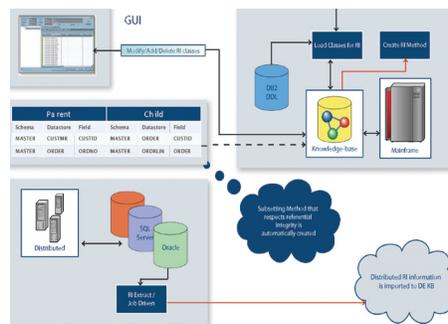


Figure 5. Cross-platform Complex Subsetting

### Rapidly Define and Create Synthetic Data without a Related Source Environment

In many cases the process of generating a test environment by subsetting and/or masking existing data is not a viable alternative. This tends to happen when new or modified tables are not

present in the source environment for test data population. For these scenarios, Data Express provides the Data Generation module. It can generate synthetic data, not related to production environment, but determined via defined, user interface rules. This module works seamlessly with other Data Express modules, using information stored in the knowledge base.

The properties of the synthetic data generation process are:

- Based on the model of the application, and is created accessing target environment structure, or an intermediate metadata format
- External key columns are populated according to the content of the parent table in the model. Parent table can already exist in target environment or be the result itself of synthetic data generation.
- Other columns can be generated according to different rules from different sources, including generation from a dictionary, from domain or random value lists, or via referential integrity.

### Comprehensive Operating System and Data Store Platform Coverage

Data Express can be used in both mainframe and non-mainframe (Windows, UNIX and Linux) platforms across a wide range of data stores.

### Data Express for z/OS

Data Express for z/OS allows the central Data Express knowledge base to reside on an IBM Mainframe z/OS environment providing a strategic central point for subsetting and masking of multiple mainframe data sources and distributed data (though the ODBC and Oracle extensions).

- **Platforms:** IBM eServer System z running OS/390 2.6 or later; or z/OS 1.1 or later. DB2 8 or later is a mandatory requirement
- **Mainframe Data Formats:** Sequential, GDG, VSAM, DB2, ADABAS, DL/I

#### Connect with Us

[OpenText CEO Mark Barrenechea's blog](#)



The supported input formats are:

- Sequential unload
- DB2 unload is supported in DSNTIAUL, REORG and UNLOAD format (IBM utilities), and in FIXED, EXTERNAL, VARIABLE and CSV (CA utilities)
- Direct access (for DB2, VSAM, GDG, sequential files)
- DB2 Image copies

#### Data Express for Distributed Systems

Data Express for distributed systems allows the central Data Express knowledge base to reside on a Windows platform removing the requirement for an IBM Mainframe. The Windows knowledge base provides a strategic central point for subsetting and masking of multiple distributed data sources through the ODBC and Oracle extensions which are included within Data Express for distributed systems.

#### Data Express ODBC Extension

The ODBC extensions to Data Express are an add-on to Data Express for z/OS and provide direct access to ODBC data stores residing on or accessible from Windows, UNIX and Linux.

#### Platforms:

- x86-64 running Windows
- x86-64 running Red Hat Linux
- x86-64 running SUSE Linux
- zSeries running SUSE Linux
- IBM System p running AIX
- Itanium running HP/UX
- SPARC running Solaris

#### Data formats:

- Data Express provides the ability to mask and subset a large number of data stores, including DB2 for z/OS iSeries and LUW, as well as Microsoft SQL Server and Oracle, Sybase, Informix.

More details on these data stores as well as others can be found in the Data Express product documentation.

Learn more at

[www.microfocus.com/data-express](http://www.microfocus.com/data-express)

[www.microfocus.com/opentext](http://www.microfocus.com/opentext)