Extracting Your Company's Data with the New Audit Data Standard

Written by Kristine Hasenstab and Eric E. Cohen

Have you ever been responsible for meeting an internal or external auditor's request for data from your accounting systems? Or does management, investor relations, or the compliance office have standing requests for information used in business modeling, reporting, and analysis? Would you like to quickly and accurately fulfill these requests? A new series of standards from the American Institute of CPAs (AICPA) for organizing general ledger and sub-ledger data may reduce the effort involved in finding, preparing, and sharing data. These new data standards are called the [Audit Data Standards (ADS)](#).

This article introduces the ADS and addresses basic implementation steps companies should consider that will help them realize benefits of the ADS.

Introducing the Audit Data Standards

On August 12, 2013, the AICPA Assurance Services Executive Committee Emerging Assurance Technologies Task Force published the first set of the ADS, which are voluntary, recommended data standards that facilitate the extraction of information from accounting systems. This article will provide guidance and insights on the value of the ADS and document the steps an organization may take to create the files for internal and external audit processes or other purposes (e.g., compliance monitoring, internal management reporting, or business intelligence (BI)).

One of the challenges that management and auditors face is efficiently obtaining accurate data from accounting systems in a usable format following a repeatable process. The ADS establishes a new, standardized way for companies to share data that is helpful to consumers of data—both auditors and management. The first set of the ADS defines key general ledger and accounts receivable file definitions, along with shared file details. When companies provide their setup, master, and transaction files in the ADS format, their auditors and management can gain accelerated insights from the data, with the promise of streamlining information exchange, improving analytics, and reducing the burden of preparing, sharing, and analyzing data.

Over time, implementation of the ADS will lead to less involvement from IT and will provide business professionals across the organization increased access to standardized information stored across a broad range of internal structured data stores. Internal auditors, investor relations, compliance office personnel, planning and budgeting personnel, tax preparers, and others within an organization can benefit from having more efficient access to this standardized information.

Basic steps of producing ADS-compliant files: an overview

The specifics steps required for creating ADS-compliant files will vary based on your accounting system(s), available report writing, export and extraction capabilities of the products you have installed, and the accounting and technical skills available to your organization. In addition, the ADS provides two specified syntaxes (technical formats) for the ADS files—pipe ("|") delimited, which we will refer to as

PipeDF, and XBRL Global Ledger Taxonomy Framework ([XBRL GL](#)). Certain preparation steps differ based upon these factors and vary based on the needs of both the internal and external users of the data.

The AICPA chose to provide both formats as an option for auditors and organizations for a number of practical and technical reasons. Data entered into fully ADS-compliant PipeDF or XBRL GL files can be transformed from one format to the other automatically. However, there are different benefits and trade-offs associated with these technical formats. The PipeDF format is relatively simple to create and view, and unzipped PipeDF files are generally smaller than unzipped XBRL GL files. XBRL GL files, in contrast, are more complex to create and review manually but with the complexity comes a wide variety of automated tooling to enhance data quality and promote reuse across a broader range of consuming applications.

Some of the steps to create ADS-compliant files are more technical in nature than others; these will likely require the involvement of your IT staff, software developer, or a consultant. Other steps will require understanding the accounting conventions in your system and how they relate to the conventions specified by the ADS. For example, the ADS for general ledger require an understanding of how different currencies—local vs. reporting currency—are captured and reported in the system. Access to your systems documentation, especially file system definitions and descriptions, will be helpful.

Management must be involved to obtain an understanding of the overall processes in order to i) obtain and oversee the appropriate resources both for initial setup and ongoing processing, and ii) ensure that there is appropriate governance over workflow (including authorization, versioning, transmission, and archival) and security for the data. This understanding will also allow management to determine if there could be additional benefits to leveraging the ADS for internal audit assessments, consolidation and aggregation, system integration, data migration, data archival, governance, risk management and compliance activities, BI, or other beneficial purposes.

The implementation steps—many of which do not need to be performed again after the initial setup (unless there are changes to the underlying systems)—are summarized below and then explained in detail.

1. Develop an understanding of the ADS files you wish to exchange with your external auditor or internal users of data

2. Identify the files/tables in your accounting system(s) necessary to obtain the data required

3. Determine the best method to extract the information

4. Map from the available data fields to the ADS fields

5. Evaluate which data transformation is necessary

6. Perform the extraction

7. Validate the resulting files

Basic steps: detail

Let's look into each of the steps in more detail. The following sections address the technical concepts that will be most helpful to those tasked with converting data into ADS-compliant data.

Develop an understanding of the ADS files you wish to exchange with your external auditor or internal users of data

The ADS defines a number of file descriptions and definitions. Depending on the work being performed, the user may want to have everything that is defined in the ADS or only selected files. For example, an auditor may only want the general ledger files for interim work and not require the accounts receivable files.

Once you have determined the requested data, you can gather the relevant file descriptions and definitions provided in the ADS. You may still wish to gather the file descriptions and definitions for all of the files in the ADS, even if all of the content has not been requested. This may prove more efficient if, for example, you are engaging external resources or anticipate that other files will be needed in the near future. Remember that certain concerns, such as privacy concerns, may require additional care when preparing to extract data.

It is important to note that the ADS are meant as a "minimum standard" and auditors and organizations can expand data being shared by adding additional fields to the end of a record when using PipeDF files and using additional XBRL GL concepts (standard or extended), where necessary.

The accumulated set of files and fields becomes the basis for the next series of steps.

Identify the files/tables in your accounting system(s) necessary to obtain the data required

The next step is to find where the requested data is located within systems and files.

Systems

The requested data may be found in multiple systems. For example, your customer data may be spread among multiple entities and multiple operational systems.

Files/tables

The operational system will be designed for security, performance, and other requirements that may lead to the requested data elements for a single ADS file (such as the *Customer_Master_YYMMDD*) being spread across multiple files or tables. Some factors to consider include:

- Most system transaction designs separate line item detail from common header detail.

- Master file detail is often normalized, meaning only a key, such as the customer's number, is included in the transactions, while the details are found in the master file itself.

- Likewise, setup files (also referred to as look-up tables), such as warehouses, tax codes, and other code sets, are simply referenced within the transaction file by a code, and the descriptions and other content is found in the setup file itself.

- Depending on the timing of the extract, historical files may be archived and the details found in separate files or tables from the active content.

In summary, you will need to understand all of the relevant files necessary to bring the content together in a single format.

Determine the best method to extract the information

Now that you know where the information is located, you can evaluate the best way to extract the data out of the system(s). The extraction decision may be driven by the nature and size of the data targeted for analysis, data security and system performance considerations. The method of extracting data may also

include transforming it into the ADS format, or there may be additional data transformation steps that are necessary to perform after the data is extracted. Examples of methods that can be used to extract the data include the following:

- Some systems permit or require *Interactive Export* of content to allow ad-hoc extraction of data. This may permit more choices on how to extract the data, however. This is not an automated method and it may not be as efficient for recurring use. This method also requires the authorization of an appropriate user on demand.

- *Batch processes* that are set up in advance can be performed automatically by schedule. If the data needs do not change other than certain filters (such as the time period requested), this method can be efficient. This method may be less efficient if the data requirements change regularly, and the batch definitions need to be altered with each data extraction.

- A *Report Writer* is more independent of the underlying system(s) and can include prompts for date and other filtering criteria. Depending on the Report Writer, multiple systems' data can be accessed and potentially combined when creating the ADS file.

- *Third Party Mapping and Extraction Tools* permit similar functionality as described in the previous methods and offer different levels of assistance in identifying and accessing the required information. Broadly known as ETL (Extract-Transform-Load) or data mapping tools, a connection between the source data and the ADS definitions can be established once and reused period to period and across similar systems. Third party mapping applications are reasonably common and can be purchased for between $200 and $1,000. ADS implementation can have a nominal impact on existing system infrastructure by using third party mapping applications.

- Many systems now offer *Web Services* access to their content. Web Services provides standardized requests and responses using XML. In a related fashion, a number of ERP vendors have special tools to create XML files from their content upon demand. XML files can be transformed from their native format to other XML or text formats using standardized transformation languages, such as the Extensible Stylesheet Language (XSL).

- As a last resort, *Spooled Reports* are available from virtually any system and software. Existing reports are printed "to disk" rather than to "paper" and are then available for manipulation. An additional step to parse the necessary content from the spooled report will be required; many data extraction and audit products include report parsing capabilities. This option involves manual processes, which can introduce a higher chance for mistakes and errors in the preparation of the information. Therefore, caution is recommended when using this approach.

Map from the available data fields to the ADS fields

Having selected the files and fields from your source systems, you need to begin the process of mapping the files and fields to the ADS fields. For the PipeDF format, you need to make sure the field content is in the order defined in the ADS. For XBRL GL, you will be mapping to the fields and structures of the XBRL GL taxonomy element as defined in the ADS.

The following graphic shows a simple example of a typical mapping tool. The left column lists the available data fields source table and the right lists the standardized fields of XBRL GL. The appropriate match between the data field source table and standard fields of XBRL GL is determined by the "XBRL GL Taxonomy Element" column of the ADS. For instance, mapping the data in the source field *Entry_Date* to

the associated XBRL GL field *gl-cor:enteredDate* is accomplished by connecting the source and target, a one-time task that is reused.
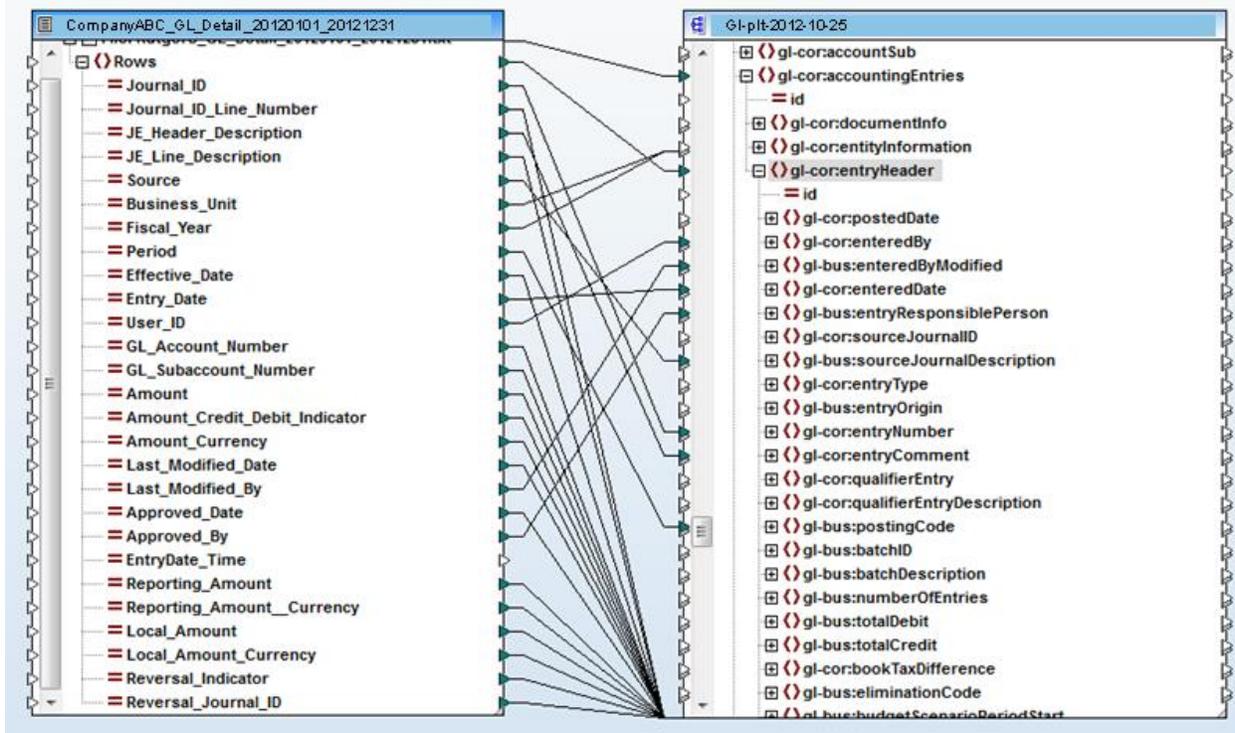
The information in the ADS may require the content from multiple fields in your underlying source table. In this case, you may need to concatenate the fields (connect separate fields together as one) in the source to create the field in the ADS. The following illustrates an example where three fields, including the document number, document type, and year, need to be concatenated into one field in order to create a unique transaction ID:

Source System				ADS
Document Number	Document Type	Year		Transaction_ID
890234	Invoice	2013	→	890234Invoice2013

Likewise, you may need to parse out content from a "combined" field in the source table in order to get that data you need for the ADS.

Another challenge in the mapping process occurs when the data comes from multiple source tables that are linked by a key, such as a customer number linking the customer master to the transaction header and the transaction header to the transaction lines. The method of joining these files together, which may involve using query techniques such as INNER JOIN and OUTER JOIN, must be carefully considered to retrieve all of the necessary components while minimizing duplicate, redundant, or missing content.

Evaluate which data transformation is necessary

The data needs to conform to the conventions, restrictions and other standards that are required by ADS PipeDF and XBRL GL formats. Different databases store numbers, dates, currencies, regions, and other

important information in different manners. When using report writers to extract content, presentation rules and formats can also obscure important attributes about the data (e.g. minus sign, data type).

For these reasons, during the extract or a later conversion step, content will need to be transformed to meet the requirements of the ADS. The transformations include:

- Text must be left-justified and trimmed (no leading or trailing blank spaces)

- Replacing characters illegal in XML (for example, &,',",<,>) with their escaped alternative (XBRL GL only)

- Converting from non-ASCII/UNICODE to ASCII/UNICODE. This is especially important when the source is an EBCIDC[1] or other non-ASCII system

- Limiting the size of text fields to those specified (PipeDF only)

- Converting dates to ISO 8601 format (CCYY-MM-DD or CCYYMMDD)

- Converting regions to ISO 3166 code formats

- Converting unit formats to ISO 4217 (currency)

- Converting numbers to their PipeDF or XML formats (e.g., no thousand separator, period as decimal point)

- Converting percentages to the specific formats (12 percent is 12.0 for PipeDF and .12 for XBRL GL)

- Mapping, where necessary or helpful, to PipeDF or XBRL GL enumerations (e.g., the Amount_Credit_Debit_Indicator in AR_Activity_YYYYMMDD_YYYYMMDD *must* be "C" for credit or "D" for debit.)

It may be necessary to create certain fields (especially Level 1 fields, which are listed as "required" in the ADS) by formula or convention. For example, a system that maintains the *document date* and the *number of days until due* may not maintain the *due date* itself; to provide the mandatory (Level 1) field Transaction_Due_Date, the date may need to be calculated based on the data in the system.

Perform the extraction

After completing the preceding steps, you are ready to perform an extraction. You will be applying certain filters (in particular, time period constraints). As noted, for the PipeDF format, the content must be in the order specified in the ADS, delimited with "|". Likewise, for XBRL GL, the substructures should be duplicated as necessary and the addition of required or helpful fields should be considered. Further, the ADS provides recommended additional XBRL GL fields that help disambiguate the export files' identification. The organization and users can easily add other agreed-upon details, already anticipated by the broader XBRL GL taxonomy, without impacting the underlying ADS or requiring change to XBRL GL.

You should agree on the naming conventions for the files. While the table names in the ADS are a useful starting point, identifying different entities, time periods and versions of extracts can be very helpful in managing the audit trail. This will be important when tracking exports and versions and determining which versions have been exchanged.

---

[1] Extended Binary Coded Decimal Interchange Code

Validate the resulting files

Before the extracted and transformed files are ready to be exchanged, you should take steps to ensure the quality of the extracted files. While the PipeDF files can be brought into Excel and manually examined or special programs can be written to make sure the text-based files meet the data needs described in the ADS (such as enforcing that dates look like dates and numbers look like numbers), the XBRL GL file format can be automatically subjected to XML, XBRL, and other agreed-upon validation and analytical tests. XML Schema validation will ensure that dates and numbers are properly formatted; additional rules, using standards like XBRL Formula or ISO Schematron, can further make sure the files conform to the requirements of the ADS.

With these seven steps in place, you are ready to begin preparing for more efficient data exchange using the AICPA's Audit Data Standard.

Kristine Hasenstab is a Director at PwC US and a CPA in Georgia. Eric E. Cohen is XBRL Global Technical Leader at PwC US and a CPA in New York. Both are active in the AICPA's Emerging Assurance Technologies Task Force Audit Data Standard Working Group.