

UDiFF Standards for Generation and Implementation

1. BACKGROUND & OBJECTIVE

1.1 There are multiple Market Infrastructure Institutions (MIIs) and segments in the Indian securities market. To exchange information among these MIIs and their members, certain files are generated by the MIIs.

1.2 The information can be broadly classified into two categories:

1.2.1 Trading and clearing operations

1.2.2 Depository operations

1.3 Currently, there are more than 200 different files generated for exchanging and clearing information by MIIs, transferred to their members, taking all MIIs and segments into account. The Market Data Advisory Committee (MDAC) observed that these formats need to be standardized and harmonized. This harmonization and standardization will facilitate the ease of doing business in the Indian securities market.

1.4 The exercise of harmonizing and standardizing the files began in 2020, initially led by the MIIs. It was decided that the files would be harmonized and standardized using ISO tags across the MIIs. During a review at the end of 2022, it was observed that there was further scope for improvement in the formats and the proposed standardized files were not fully meeting the requirements of the MDAC. SEBI took a holistic approach, gathering feedback from end-users and later informing MDAC of their suggestions.

1.5 The task of further standardizing and harmonizing the files was entrusted to the SSU & MDAC division of the Department of Economic and Policy Analysis of SEBI in June 2023.

1.6 Broadening of the Scope of Implementation –

1.6.1 A detailed study of the files was conducted SEBI after assuming the responsibility of publishing catalogue. In order to user centric catalogue, end-users were first consulted by the division. It was concluded that to harmonize and standardize the files, rationalization of the files and fields (variables) was necessary to resolve multiple legacy

issues observed by the SEBI team. Further, to simplify IT development and future maintenance, discussions with stakeholders revealed that the number of files could be minimized.

1.6.2 In-depth discussions with market participants indicated that the over 200 different files from the exchanges and CCs provide 11 different types of information to undertake 11 different tasks. Accordingly the approach was changed from the file based standardization to information based one. As a result, more than 200 different files could be unified to 11 simpler file formats. The fields in each of the formats were also distilled to keep only rationally required fields. Additionally, a few more fields as suggested by the market participant were added in order to facilitate them in development and processing. The project, with its broadened scope, was named Unified Distilled File Formats (UDiFF).

1.7 Rationale -

1.7.1 A single set of information across segments (equity, currency derivatives, equity derivatives and Commodity derivatives) has a large number of common fields (variables). Only a few fields are added or removed depending on the segment. Providing different files that require different processing logic increases the cost of development and subsequent modifications. Simplifying the process to the essential minimum leads to a reduction in the cost of doing business.

1.7.2 In UDiFF, one type of information was made available in a consistent format across segments and institutions. A direct advantage of UDiFF thus envisaged was the consolidation of development and future maintenance to a single format, instead of close to 10 different formats as was the case earlier. UDiFF formats are also expected to simplify current development for standardization and harmonization exercise along with streamlining of future changes.

1.7.3 Take, for instance, trade information, which contains details of executed trades. It's one of the most frequently used formats, downloaded several times daily for various requirements. For nine MIs (five exchanges & four clearing corporations) across four segments prior to UDiFF, there would be 28 distinct trading file formats. Moreover, trading and clearing members would have different file formats, effectively doubling the number

to 56. In some cases, the same file for the same segment and MII would differ when downloaded from FTP. As mentioned earlier, many fields in these trading files would be identical.

OUTCOME:

1.7.4 More than 90% of the cost reduction in a few formats -

1.7.4.1 Trade information format - A UDiFF format for trade files was created after consolidating over 40 trade files comprising 112 unique fields. All these trade files were "unified" into a single format, and the 112 unique fields were "distilled" down to fewer than 40 fields.

1.7.4.2 To achieve this, each field was rigorously assessed. To ensure alignment with end-users' needs, market participants were continually consulted throughout the process. In addition to the new practice of ongoing user feedback, the entire catalogue was also proposed to be published in the public domain, echoing SEBI's ongoing efforts to increase transparency, democratize data access, and reduce business costs.

2. Design of UDiFF –

2.1 UDiFF formats are designed to encapsulate one type of information within a single format type. The aim was to include all fields in a single file comprising one table to process that information. Given the vast amount of information generated, CSV was selected as the communication format, incorporating ISO tags as per ISO 20022 for securities market messages (Financial Services) and ISO 8601 for date & time-related fields. Furthermore, UDiFF was envisioned to serve as a consolidated reference document. Thus, a user referring to UDiFF wouldn't need to consult relevant circulars issued by MIIs. Summaries of changes will also be provided within UDiFF, detailing implementation dates and sources (such as the circular) of respective changes. This information is exclusive to changes in UDiFF and is expected to function as a log file for the entire system.

2.2 About the UDiFF Catalog –

2.2.1 The catalogue is available in Excel format and is organized into seven sheets detailing information related to exchange, clearing and settlement operations:

2.2.1.1 **Principles** - These help end users including developers using the formats to process information, be it generation or consumption. These principles are derived from those listed in this document, primarily targeting development within that format.

2.2.1.2 **Focused Format-wise Field List** - This is a streamlined, format-specific list optimized for easy reading, leveraging the advanced filtering features of MS Excel.

2.2.1.3 **Format Master** - This documents all formats present in the current UDiFF version.

2.2.1.4 **Field Master** - This documents all fields, organized by format.

2.2.1.5 **Standard Value List** - This lists fixed values for specific files.

2.2.1.6 **Change Log** - Updated based on feedback and highlights changes in the current version relative to the previous iteration.

2.2.1.7 **Open for Comments** - As feedback is accumulated and discussions ensue, this sheet will serve as a central repository for open suggestions and concerns. Given that various entities from MIs will use it, and not everyone will have email access to current discussions, this approach ensures transparency and accessibility for all.

2.3 Structure of UDiFF Formats

2.3.1 The format: One file will have one table only. All headers will be in the first row, and data values will start from the second row onwards. It will not be permissible to append another table directly after one table or after leaving a few rows.

2.3.2 All files will have ISO headers in their first row for the identification of the columns and easier readability of the files. The header and order of the columns in the CSV will be adopted from the catalogue file.

2.3.3 The values, wherever prescribed by ISO standards (ISO 20022), will also conform to the ISO standards.

2.3.4 There will be no entity-specific fields, and all values are to be populated by all MIs. Values will not be populated only in cases where they are not applicable. For example,

for a cash market trade file, these fields can be derivative fields. Otherwise, all fields shall be mandatorily populated by all entities.

2.3.5 Multiple files can be generated for the same format such as a trade file can be downloaded separately for cash, equity derivative, and commodity segments. However, the format of all those files is going to be identical as prescribed in this UDiFF catalogue.

2.3.6 Amounts will always be in rupees and up to 6 places after the decimal.

2.3.7 Dates and time will be in ISO 8601 / IS 7900 (2007-03-01T13:00:00) (YYYY-MM-DDTHH:mm:ss) format. In case time is also applicable, then it will be part of the same Date column following the ISO 8601 standard. In case time is not applicable the format shall be YYYY-MM-DD.

2.3.8 Files will be in CSV format exclusively, and the fields will be separated by a comma. The Encoding of the files would be **ASCII**, and the line-ending character should be `\r\n`. This is the same being used by windows and the majority of the other MIIs in existing systems. For compression, .gz or .zip will be used for large files.

2.3.8 Standard values inside a field will be used as the content, as provided in the “Standard Value List” sheet of the catalogue. Standard Values would be character-based to make them more readable instead of numbers, similar to the principles adopted by ISO during the change from ISO 15022 to ISO 20022.

2.3.9 ISO tags and ISO values will be used where available.

2.3.10 In case of conflict among MIIs, for example, if one MII has a field length of 10 characters and another has 12 characters for the same field, the 12-character length will be adopted. It is understood that a larger field size can accommodate smaller values as well. Thus, filler characters need not be there.

2.3.11 Unique Code for Each Field

2.3.11.1 The field list will have a unique field identifier (already included a column for the same); codes have been generated using the name of format name and clubbing the same with a three-digit number, starting with 101. For example, the Trade file's first field will have the code "Trade101", "Trade102".

Rationale -

2.3.11.2 It was observed that many times sorting of the fields needs to be done, and there having this structure helps in maintaining the correct order, otherwise "trade1" is followed by "trade11" instead of "trade2".

2.3.12 Dummy Fields and Remarks Field

2.3.12.1 All formats will have a remarks field Varchar(150) followed by 4 dummy fields of varchar(50). The rationale for having these has already been provided in this document's latter part.

Rationale -

2.3.12.2 Rationale for the additional fields and remarks field has been mentioned in "**Need based changes and Periodic Review**" section.

2.4 Nomenclature of the files

2.4.1 The name of the files will consist of:

2.4.1.1 name of the file as per format name (Format Master) of the UDiff Catalogue,

2.4.1.2 followed by the 10 character code of the entity generating it,

2.4.1.3 followed by a two-character segment code (or 0 if not applicable),

2.4.1.4 Followed by Market (or 0 if not applicable). The market shall be EquityT1, EquityT0, Auction etc.

2.4.1.5 followed by TM or CM (or 0 if not applicable)

2.4.1.6 followed by a unique trading or clearing member code (or 0 if not applicable),

2.4.1.7 the date in “YYYYMMDD” format. If a file is applicable for a month the for e.g. Settlement Master the date shall be populated as 20231100. The day DD shall be printed as 00.

2.4.1.8 the variable for indicating if it is a provisional or final report “P” or “F”

2.4.1.9 and finally “hhmm” which will be in a 24-hour format or preceded “0000” (Zeros) when the file is final file.

Example: if a file is generated at 7 minutes past 9 AM. It will be “0907” appended in the names. All of these values will be in the same order as described above and will be separated by “_” (underscore) character. Accordingly, the underscore character will be used in the format or file names or format names will only be in Pascal Case¹. In case of Intraday margin file which are generated multiple times, as per regulatory guidelines, the time stamp shall be 1100, 1230, 1430, 1530, 1700, 1900, 2030, 2230 etc. and 0000 for EOD. The timestamp in Margin file nomenclature shall indicative of the sequence of generation of files; it is not indicative of the time at which the data snapshot has been taken for generation of the files.

2.4.1.10 The entire name string for the example trade file generated for the example given previously can be referred to as the following:

For Provisional File:

Exchange

Trade_BSE_CD_0_TM_<<member_code>>_20230601_P_0000.csv

BhavCopy_BSE_CD_0_0_0_20230601_P_0000.csv

¹ **Pascal Case** (“ThisIsPascalCase”) | **Camel Case** (“thisIsCamelCase”) | **Snake Case** (“this_is_snake_case”) | **Kebab Case** (“this-is-kebab-case”)

Clearing Corporation

Obligation_NCL_CM_Auction_CM_<Mem code>_YYYYMMDD_P_0000.csv

Obligation_ICCL_CM_Auction_CM_<Mem code>_YYYYMMDD_P_0000.csv

For Final File:

Exchange

Trade_BSE_CD_0_TM_<<member_code>>_20230601_F_0000.csv

BhavCopy_BSE_CD_0_0_0_20230601_F_0000.csv

Clearing Corporation

Obligation_NCL_CM_Auction_CM_<Mem code>_YYYYMMDD_F_0000.csv

Obligation_ICCL_CM_Auction_CM_<Mem code>_YYYYMMDD_F_0000.csv

Intraday Margin files

Margin_NCL_FO_0_TM_<memcode>_<YYYYMMDD>_P_hhmm.csv

Margin_ICCL_FO_0_TM_<memcode>_<YYYYMMDD>_P_hhmm.csv

2.4.1.11 Rationale -

In order to be processed by a software without exception, the file names (csv files) being generated from the system need to be standard with equal number of underscore characters in every scenario. The underscore was proposed initially as a separator for format names such as “Bhav_Copy”. This is likely to cause issue now as the finale name string (for CSV files) also using it as a separator. File name accepts only a few special characters, in order to be used in CSV file names, the underscore is being removed from the individual format (or file) names as mentioned in the UDiFF. Accordingly, **Bhav Copy** file, which was earlier proposed to be written as “Bhav_Copy” would be now written as

“BhavCopy” in Pascal Case. Pascal Case will be adopted for all formats/files names being included in the standard in “Format Master” and “Field Master” sheets of the catalogue.

2.4.2 The hours and minutes being added will be the time of the request or the time of the files being ready or was downloaded in the file name or the time as per regulatory guidelines.

2.4.3 In case there is only one file or some file is the final file such as a bhav copy/margin files in case of multiple sessions during the day. The hour and minutes fields will carry only “00” so the bhav copy for the cash segment at BSE would be, as the member code is not applicable for the same and it is common for all:

Exchange

BhavCopy_BSE_CD_0_0_0_20230601_F_0000.csv

Or

Clearing corporation

Margin_NCL_FO_0_CM_<memcode>_<Trade date>_F_0000.csv

Margin_ICCL_FO_0_CM_<memcode>_<Trade date>_F_0000.csv

3. Downloading

3.1 Functionality of Part Download to All -

3.1.1 In case of large files, a few of the MII members are facilitated to download the files in parts. It is understood that this functionality is made available only through a mechanism of request. This functionality is very useful for others as well as learned from the feedback and should be made available through the portal itself. Any member should

be able to enable and disable this functionality from their logged-in page. For a standardized process, a cap in the minimum number of rows each file can be fixed, such as 1 crore (10 million rows). Where MII are downloading files to member interface as part of push mechanism, they may provide multiple file once such cap is reached.

3.1.2 Rationale –

3.1.2.1 It should make the process smoother for medium to large members and users, and a few of them may also not be aware of the fact that such functionality exists but may or may not be available to them currently. This follows the principle of equal access.

3.2 Order of download –

3.2.1 The parts files will be allowed to be downloaded sequentially in order to load them into the system with the minimal possible delay. The file being downloaded in parts will follow the described nomenclature. In addition to identification, it will have “_p1”, “_p2” appended to it. However, since one would not know which is the final part, the final part should have “_p0” affixed to it.

Illustration:

For Exchange

Sample Splitted Provisional file name :

Trade_NSE_CD_0_TM_<<member_code>>_20230601_P_0000_P1.csv.gz

Trade_NSE_CD_0_TM_<<member_code>>_20230601_P_0000_P2.csv.gz

Trade_NSE_CD_0_TM_<<member_code>>_20230601_P_0000_P0.csv.gz

Sample Splitted Final file name :

Trade_NSE_CD_0_TM_<<member_code>>_20230601_F_0000_P1.csv.gz

Trade_NSE_CD_0_TM_<<member_code>>_20230601_F_0000_P2.csv.gz

Trade_NSE_CD_0_TM_<<member_code>>_20230601_F_0000_P0.csv.gz

For Clearing Corporation

Delivery_NCL_CM_Auction_CM_<Mem code>_YYYYMMDD_F_0000_P1.csv

Delivery_ICCL_CM_Auction_CM_<Mem code>_YYYYMMDD_F_0000_P2.csv

3.2.2 Rationale –

3.2.2.1 A few files such as the statement of holding take around 45 minutes for generation and final download as observed during the one of discussions. It is shared as an archive file with the entity using it. Assuming it contains 9 parts and each is taking approximately 5 minutes to generate. Allowing the part files to be downloaded and processed would enable the system to start consuming the file from the 6th minute instead of the 46th minute, assuming the download time is negligible.

3.3 Public Availability of generic non confidential files

3.3.1 It has come to our attention that several files, which are accessible through secured means, are generic and don't contain any confidential information. These files are the same for every entity downloading and using them. One such file is the "script master" or "contract master"; this file is meant to help systems initiate trading by detailing securities available for trade on the exchange for a particular day. Another example is the "corporate actions" file, which provides information about events that occurred during the day. If an entity doesn't have exposure to a particular script, that script is not made available to them.

3.3.2 It is recommended that these files be made publicly available on the publishing entity's website, covering all pertinent details.

3.3.3 Rationale –

3.3.3.1 There are occasions when files might not be used on the day they're provided but are still needed by the entities. If they don't have official access to the file, they might resort to unofficial means to obtain the same information, which is not desired. Moreover,

SEBI has always envisioned democratizing data. Putting this non-confidential data in the public domain would allow many researchers to access a wealth of information, offering synergistic effects when combined with existing data.

4. Nodal Official from the MIIs

4.1 Every MII will nominate a nodal person or designate a group email ID. This contact will be for those who wish to raise queries or voice concerns about the formats.

4.2 Rationale –

4.2.1 While it's understood that existing members have access to contact information, it may be challenging for newcomers to connect with the appropriate contact. For instance, a startup creating a new system might struggle to reach someone familiar with the UDiFF formats.

5. Publication of the UDiFF & Testing Files on Respective Websites

5.1 UDiFF catalogue along with this guide will be made available by every MII involved on their own website by way link on respective website or issuance of circular. The page will also list the nodal official's name or group email ID for inquiries and testing files as detailed in next sub-section.

5.2 Link to the SEBI website for the latest standards and developments. MIIs will check the SEBI website monthly and update the latest UDiFF on their own websites when there are updates.

5.3 Rationale –

5.2.1 Historically, these formats were not publicly accessible. This restricts an individual's or an entity's ability to prototype for the market. Entities trying to address a challenge need to estimate the required resources for a cost-benefit analysis. They need

definitive answers to questions like: "Is solving this problem viable as a business?" These formats not only reduce entry barriers by lowering the business costs but also enhance transparency.

5.4 Testing Files

5.4.1 The UDiFF catalogue will come with test files. These will cover every potential scenario for each format and will be provided by the respective MII, accompanied by a processed output file, both before initial development and for any subsequent format changes. These files won't contain confidential information but will follow all the logic of system-generated files. For confidentiality, data can be masked with structurally similar fields. For instance, an original PAN might be replaced with a 10-character alphanumeric string that isn't an actual PAN.

5.4.2 Each test file should include several rows for every possible scenario, enabling development teams to test their implementation.

5.4.3 Rationale –

5.4.3.1 The entire initiative aims to enhance accuracy and efficiency while also reducing system costs. Given that most formats are generated by the MIIs, their development is time-consuming. The advantages of these test files are twofold.

5.4.3.2 Firstly, entities dependent on these formats can develop and test their systems during the development phase rather than waiting for the MIIs to finalize their parts.

5.4.3.3 Secondly, the MII development team can test their implementation using these files. The benefits aren't confined to MIIs' internal use. It enables other entities implementing the formats to offer comprehensive feedback to the MIIs if there's an issue or if another MII implements it differently and there are incorrect logical differences.

5.4.3.4 In conclusion, this approach shortens the development cycle and promotes a smoother, more efficient, accurate, and cost-effective development process.

6. Need based changes and Periodic Review

6.1 Feedback Collection and Deliberations Before Changes

6.1.1 The formats will be reviewed as needed based on requests from any stakeholder or in response to regulatory or market developments. To prevent desynchronization, every format change will be implemented only after comprehensive feedback from all involved parties. Feedback will be collected at the SEBI/concerned MDAC working group level. After deliberation, the most suitable method for implementing the change will be incorporated into UDiFF and communicated appropriately.

6.1.2 Rationale –

6.1.2.1 Representing the interests of the entire market and balancing stakeholder expectations and implementation feasibility requires an independent nodal point. This nodal point aids in identifying optimal solutions that address all stakeholders' needs while minimizing overall costs for the entire ecosystem.

6.2 Changes for New Products or Offerings by Select MIIs:

6.2.1 For situations where changes are necessary for some MIIs but not all, these changes will preferably utilize the dummy fields in each format. However, consultations with MDAC & SEBI will take place, and only agreed-upon changes will be incorporated into the formats after publishing and notifying stakeholders through UDiFF. The specified dummy field, out of four dummy fields available in each format, will be reserved by MDAC/SEBI. If another MII wants to introduce new information simultaneously, a different field will be used. It's worth noting that this arrangement applies to fields added permanently. For temporary information, the provided remarks field at the end of each format can be used without MDAC/SEBI intervention.

6.2.2 Rationale –

6.2.2.1 If an MII wishes to introduce a new product requiring additional information, facilitating this process is crucial. To foster market innovation, this approach prevents implementation delays at one MII by eliminating reliance on other MIIs. It also aims to reduce implementation costs by minimizing ad-hoc development.

6.3 Synchronization Cycle

6.3.1 Annually or semi-annually, MIIs will review the latest UDiFF catalogue and add new fields used as dummy fields by MIIs. If certain MIIs lack data for the dummy field, they can leave it blank. At this stage, all discontinued fields will be removed. The dummy fields will be assigned appropriate names, ISO tags, field codes, and detailed definitions by the entity that first implemented them before the synchronization date, enabling other MIIs to integrate them into their formats. These fields will be made available for future use if needed. At the conclusion of each synchronization cycle, the formats across all MIIs will be identical.

6.3.2 Rationale-

6.3.2.1 Dummy fields should be reserved only temporarily. Maintaining format synchronization and ensuring a sufficient number of available dummy fields is essential. Therefore, a fixed cycle where all MIIs incorporate changes and start publishing is necessary. This also helps MII members consuming these formats to anticipate and adjust to changes with proper notice.

6.4 Preferred Order of change in the UDiFF Catalogue

6.4.1 The methodology suggested here is merely guiding. Therefore, when changes are anticipated, the following method will be considered for their implementation. However, all changes will be discussed and evaluated by the concerned working group of MDAC/SEBI.

6.4.2 The suggested order of changes is as follows. It should be determined if the additional information can be conveyed by making changes in one or a combination of the following:

6.4.2.1 Adjusting the list of standard values

6.4.2.2 Adding a new field in a format

6.4.2.3 Introducing a new format or altering the logic

6.4.2 Rationale –

6.4.2.1 Every alteration has associated cost implications. In our ongoing effort to minimize costs, the objective here is to provide a roadmap to ensure minimal expenditure for the entire ecosystem when proposing changes. The aforementioned order of changes represents the anticipated ascending order of associated costs.