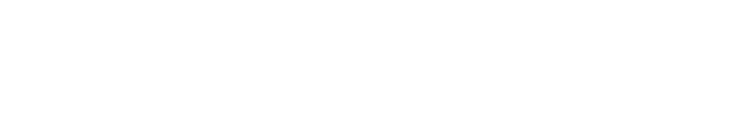
**Investment Division Data Lake**

Business Requirements Document



DataPhile to IDDL Data Feed

Data Ingestion v.04a

****

# *Document Revisions*

|  |  |  |
| --- | --- | --- |
| *Date* | *Version Number* | *Document Changes* |
| *11/15/2019* | *0.1* | *Initial Draft* |
| *11/18/2019* | *0.2* | *Workshop* |
| *12/12/2019* | *0.2a* | *IDDL changes.* |
| *01/07/2020* | *0.2b* | *Finalize pulls* |
| *2/21/2020* | *0.2c* | *remove all masking of PII fields and don’t load any columns into IDDL. , added new extract logic.* |
| *4/15/2020* | *.3* | *Clarification for design* |
| *4/21/2020* | *.4* | *Clarification for QA , added keys for three table*   |  |  | | --- | --- | | **MFCL** | Primary key: **client** | | **MFRRUS** | usr, rr | | **GNCODE** | Table\_name, code | |
| *5/8/2020* | *V.5* | *Ryan: ALM #8888 – Appendix 8*  *Added three column in Data Dictionary for MFRR to match IDDL:share\_office,other\_business,other\_business\_type* |
|  |  |  |

Contents

[*1* *Document Revisions* 2](#_Toc27638893)

[2 Project Summary 4](#_Toc27638894)

[2.1 Background & Overview 4](#_Toc27638895)

[2.2 Approach 4](#_Toc27638896)

[2.3 Project Scope 5](#_Toc27638897)

[2.3.1 In Scope 5](#_Toc27638898)

[2.3.2 Assumptions 5](#_Toc27638899)

[2.3.3 Constraints 5](#_Toc27638900)

[2.3.4 Dependencies 5](#_Toc27638901)

[2.3.5 Risks 5](#_Toc27638902)

[3 Business Process Overview / Context Diagram / System Diagram 6](#_Toc27638903)

[3.1 Current Business Process (As-Is) 6](#_Toc27638904)

[3.2 Proposed Business Process (To-Be) 6](#_Toc27638905)

[4 Business Requirements 7](#_Toc27638906)

[4.1 Business Requirements / User Stories 7](#_Toc27638907)

[4.2 Dataphile table list (section will be completed in Phase 2) 9](#_Toc27638908)

[5 Latest Version see SharePoint Non-Functional and Technical Requirements 10](#_Toc27638909)

[6 Business Rules 10](#_Toc27638910)

[7 Approvals 10](#_Toc27638911)

[8 Appendices 10](#_Toc27638912)

[8.1.1 Data dictionaries 10](#_Toc27638913)

[. See attached in sharepoint 10](#_Toc27638914)

[8.1.2 DSA 10](#_Toc27638915)

[8.1.3 Data discussion points 11](#_Toc27638916)

# Project Summary

## Background & Overview

This project is part of the enterprise wide Global Wealth and Asset Management (GWAM) Assets Under Management (AUM) project. The purpose of this project is to provide daily automated self-service reporting and analysis for WAM Finance with improved insights to business partners for better decision-making. This aligns with the GWAM Data Strategy and Technologies to provide accessible, secure and timely financial (GL) and non-financial (e.g. AUM/flows) data in aggregate and at the level of detail needed to satisfy WAM Finance reporting and analysis requirements.

The data team will provide finance end users a holistic view of multiple lines of business under one central location, Investment Division Data Lake (a.k.a IDDL).

There are multiple systems and data sources required in IDDL for this project and this requirement document is focusing on data required from DataPhile Progress database into IDDL ingestion layer. The current DataPhile Progress database contains over 20 years of history but has been used for reporting in the last two years.

DataPhile is an electronic record keeping system that services the Canadian retail business unit called Manulife Securities. Manulife Securities has $39 billion\* in assets under administration and is a wholly-owned subsidiary of Manulife Financial Manulife. The company services client with both Manulife funds as well as other types of Securities that are reffred to as Assets under administration (ASA) .

Dataphile is four Progress database with a large number of tables based on different companies :

* MSI
* MSIS
* MSII
* Securities (called VM)

Being a electronic record keeping system the system includes Personally identifiable information (PII) data which will not be part of this project.

The purpose of this BRD is the ingestion of the DataPhile system into the IDDL with a subsequent BRD required for curation. All references to reporting and curation in this document are for planning purposes and to provide designer context.

## Approach

The project approach will take two phases :

1. Land Dataphile tables with sufficient sample for data analysis in dl
2. Based on analysis create business rules required to create Business rules to calculate AUm and fund flows.

The requirements and scope will separated into each phase.As of Mid January Phase 1 has been completed.

## Project Scope

### In Scope

* Phase 1
  + Ingestion of subset DataPhile tables required for AUM with no Personally identifiable information (PII) that could be sourced from :
    - Operational Progress database
  + Secured connection between DataPhile systems and IDDL
  + Server and connection information is in scope and provided in a separated document
* Phase 2
  + Ingestion of 2 of years of DataPhile history for tables required for AUM with no Personally identifiable information (PII) that could be sourced from :
    - Operational Progress database
  + Process extract deltas from DataPhile as soon as daily refresh is completed and data are available for consumption
  + Ability to store history for each day for traceability
  + Daily load of all tables into IDDL raw zone
    - Full load when go live
    - subsequent daily loads of new data as well as mofications to previous days

Out of Scope

* Data outside of DataPhile unless mentioned as in scope
* Ingestion of near real time data
* Row level – validations and error checking
* Data cleansing and standardization

### Assumptions

* DataPhile tables are refreshed daily and data is only loaded once into IDDL at eod
* Data in DataPhile has been validated by business team through analysis conducted in Phase 1 before automating the daily loading of data into IDDL
* Data coverage in DataPhile is enough for downstream consumers
* An automated secured file transfer protocol (SFTP) process is available for DataPhile systems to support the transfer of data files through Manulife’s network infrastructure?
* DataPhile FUND identifiers’ will be mapped to IDDL Portfolio identifiers in a separate document. These Identifiers are unique codes to identify Fund in Manulife’s accounting system. The IDDL is joined and built using these identifiers. A separate group will maintain this variables and translation.
* A business rules on extracting data from DataPhile will be available to complete curation BRD with a subsequent model.

### Constraints

### Dependencies

### Risks

# Business Process Overview / Context Diagram / System Diagram

## Current Business Process (As-Is)

Currently AUM reporting for Manulife Securities is prepared through the Progress canned reports as well as extracts to the sql\_ms\_ProClarity\_rP1 database.

## Proposed Business Process (To-Be)

Proposed AUM reporting for Manulife Securities will be sourced from DataPhile to IDDL .



# Business Requirements

The requirements in this document are prioritized as follows:

|  |  |
| --- | --- |
| **Rating** | **Description** |
| Critical | This requirement is critical to the success of the project. The project will not be possible without this requirement. |
| High | This requirement is high priority, but the project can be implemented at a bare minimum without this requirement. |
| Medium | This requirement is somewhat important, as it provides some value, but the project can proceed without it. |
| Low | This is a low priority requirement or a “nice to have” feature, if time and cost allow it. |
| Future | This requirement is out of scope for this project and has been included here for a possible future release. |

## Business Requirements / User Stories

| **Phase** | **Req #** | **Priority** | **Complexity** | **Description** | **Notes** |
| --- | --- | --- | --- | --- | --- |
| One | BR-001 | High | Medium | The data in raw zone should be an exact replication of DataPhile data. There should not be any transformation, omission or addition of data during transfer process. The data in IDDL raw zone should be the same as data in source tables.  The typed tables can follow the naming convention in IDDL with data type applied to the data based on data dictionary. |  |
| One | BR-001.1 | High | Medium | The system will exclude all PII identified fields when loading the data into the in IDDL . | See [Data Dictionary](#_Data_tables_layouts) |
| One | BR-001.1 | High | Medium | IDDL must ingest from DataPhile approximately 16 tables with an initial load for two years and then daily historical updates.  Note If tables are large then a subset can be used for phase1.  Project schedule will indicate when 2 years will start. | See section 4.2 |
| Two | BR-001.2 | High | Medium | IDDL must ingest from DataPhile approximately 5 tables with subset of daily historical loads of 1 days   1. Brokerage Accounting Transaction File (actran) thru PROC\_DT 2. Contract file (tpcont) thru PROC\_DT 3. Keypunched trailers file (tpkptl) thru LAST\_DT 4. Exchange rates master file (gnexch) thru LAST\_DT 5. User codes mfrrus thru PROC\_DT   For all subsetted tables use required dates > current date minus 1 business day  For Example   * Monday you would select Friday * Tuesday you would select Monday   This requirement will be detailed in design based on table sizes. | See section 4.2 |
| Two | BR-001.3 | High | Medium | IDDL must ingest from DataPhile for approximately 10 tables with full dataloads since the table size does not warrant delta’s.   1. gngnco 2. mfclac 3. mfacct 4. mfrr 5. esmgr 6. acpos 7. mfclcl 8. mfsc 9. mfbr 10. gncode   This requirement will be detailed in design based on table sizes. | See section 4.2 |
| Two | BR-002 | High | Medium | **Date Transfer Frequency:**  A complete DataPhile table load must occur for the previous day around 5am EST. It’s assumed if the extraction process fails the table will be sent later that day when system becomes available. There is no requirement for intraday data refresh from source. IDDL needs only to load data once a day. |  |
| Two | BR-002.1 | High | Medium | **Date Transfer notification.**  The DataPhile can notify IDDL process that DataPhile is ready to be extracted and loaded through the gncode table.  Design discussion : See email in appendix | BR-002 |
| Two | BR-003 | High | Simple | There should be a control log or registry table that maintains time when data is received in IDDL. |  |
| Two | BR-004 | High | Medium | The IDDL must be able to ingest the data from DataPhile tables when daily refresh is completed from source with minimum delay. |  |
| Two | BR-005 | High | Medium | The file process/ingestion timestamp should be available in the raw zone |  |
| Two | BR-006 | High | Medium | ~~IDDL will validate the fields in each table with its metadata file and ingestion process will fail if the columns do not match with data dictionary.~~  For any changes in file structure (i.e. adding/removing a field), business will inform IDDL with updated data dictionary in advance to ensure accuracy of ingested data. |  |
| Two | BR-007 | High | Medium | IDDL should send notification to related party when the ingestion process does not start by an agreed time (TBD). |  |
| Two | BR-08 | High | Medium | IDDL should send notification to related party when the ingestion process fails, and error log is available to support team for investigation. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## Dataphile table list (section will be completed in Phase 2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | System  (new) | Dataphile Table Name | description (scraped from metadata) | Notes |
| 1 | MSI  MSIS  MSII | actran | Brokerage Accounting Transaction File - a record is created     for each trade, cheque, journal entry, etc.  Segregation        processing does not create transactions. | fact - transactions table |
| 2 | MSI  MSIS  MSII | gngnco | Company information master file | dim - list of companies (MSI, MSISI, MSII) |
| 3 | MSI  MSIS  MSII | mfclac | Client account record.  There may be several accounts for each client master record (mfcl) | accounts table |
| 4 | MSI  MSIS  MSII | mfacct | Special Account Numbers | exclude these accounts (box accounts) |
| 5 | MSI  MSIS  MSII | mfrr | RR Master File | rr = advisor number |
| 6 | MSI  MSIS  MSII | esmgr | Company Fund Name | source for Company Fund Name |
| 7 | MSI  MSIS  MSII | acpos | Brokerage Accounting Position File - contains a record for      every account's security position - pending settlement,         current and fail, segregated.  Also maintains a memo position   for portfolio (non accounting postions). Note this table contains no history. | positions - used to calculate AUA |
| 8 | MSI  MSIS  MSII | gnexch | Exchange rates master  -- pretty primitive for now! | exchange rate used for US funds |
| 9 | MSI  MSIS  MSII | tpkptl | Keypunched trailers - one record per trailer | used to identify SWITCH |
| 10 | VM | mfclcl | Account Classes | account classes - also used to exclude box accounts |
| 11 | VM | mfsc | Securities table | securities table - one record per cusip |
| 12 | MSI  MSIS  MSII | tpcont | Contract file - 1 record for each side of a trade (ex. clrg) | used to calculate net/gross value |
| 13 | MSI  MSIS  MSII | mfbr | Branch Master File | dim - branches |
| 14 | MSI  MSIS  MSII | mfcl | Client file |  |
| 15 | MSI  MSIS  MSII | mfrrus | Investor advisor |  |
| 16 | MSI  MSIS  MSII | gncode | Table which tells you when data can be extracted and loaded |  |

## Dataphile table list (with sizes and fields) based on Phase 1 analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S. No.** | **Table name** | **Table name in IDDL** | **Extraction date range** | **Range column** | **Row count** | **PII** |
| 1 | acpos | inv\_aum\_temp.dataphile\_acpos | All data |  | 1386992 | No |
| 2 | actran | inv\_aum\_temp.dataphile\_actran | year 2018 to present | PROC\_DT | 30250157 | No |
| 3 | esmgr | inv\_aum\_temp.dataphile\_esmgr | All data |  | 12 | No |
| 4 | gncode | inv\_aum\_temp.dataphile\_gncode | All data |  | 507688 | No |
| 5 | gnexch | inv\_aum\_temp.dataphile\_gnexch | year 2018 to present | LAST\_DT | 304 | No |
| 6 | gngnco | inv\_aum\_temp.dataphile\_gngnco | All data |  | 2 | No |
| 7 | mfacct | inv\_aum\_temp.dataphile\_mfacct | All data |  | 220 | No |
| 8 | mfbr | inv\_aum\_temp.dataphile\_mfbr | All data |  | 471 | No |
| 9 | mfcl | inv\_aum\_temp.dataphile\_mfcl | All data |  | 419306 | Yes |
| 10 | mfclac | inv\_aum\_temp.dataphile\_mfclac | All data |  | 732286 | Yes |
| 11 | mfclcl | inv\_aum\_temp.dataphile\_mfclcl | All data |  | 94 | No |
| 12 | mfrr | inv\_aum\_temp.dataphile\_mfrr | All data |  | 9324 | No |
| 13 | mfrrus | inv\_aum\_temp.dataphile\_mfrrus | All data |  | 29700 | No |
| 14 | mfsc | inv\_aum\_temp.dataphile\_mfsc | All data |  | 1957501 | No |
| 15 | tpcont | inv\_aum\_temp.dataphile\_tpcont | year 2018 to present | PROC\_DT | 4713026 | No |
| 16 | tpkptl | inv\_aum\_temp.dataphile\_tpkptl | year 2018 to present | LAST\_DT | 8580755 | No |

# Latest Version see SharePoint Non-Functional and Technical Requirements

|  |  |  |
| --- | --- | --- |
| **Phase** | **ID** | **Requirement** |
| One | NFR-001 | PII data cannot be loaded into IDDL |
| One | NFR-002 | Access to DataPhile data in IDDL should be restricted to authorized security profiles |
| Two | NFR-003 | IDDL must be able to ingest DataPhile data upon trigger and trigger the IDDL curation when all databases have been loaded. |
| Two | NFR-004 | Data ingested in IDDL will be retained indefinitely . History cannot be deleted. |
|  |  |  |

# Business Rules

No business rules are applicable for this BRD, as it documents ingestion of data from DataPhile to the IDDL Raw Zone without any transformations with exception that all PII fields are not included.

# Approvals

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Role** | **Approvals** | **Contributors** | **Distribution** | **Name** | **Title** |
| Project solution |  | x | x | Alexandra Dascalu | Solutions Architect |
| IDDL Project Manager |  | x | x | Harneet Aulak | Project Manager, GSD |
| Development Lead |  |  | x | Krithika Swaminathan  Daniel Chang | Technical Lead, GSD |
| Data Governance |  | x |  | Tetsuji Maniwa | Director - GO Data Design |
| Data Governance | x |  | x | Todd Warren | AVP Data Design |
| WAMFIT finance |  | x | x | Nishant Mohan | Manager, Finance Transformation: Global Wealth and Asset Management |
| WAMFIT finance |  | x | x | Dino Basic | Manager Reporting & Analysis |
| WAMFIT finance | x |  | x | Josh Kester | AVP Process Improvement |
| DataPhile Owner | X | X | X | Doug Tyson | AVP Application Support, Global Wealth & Asset Management Technology |
| DataPhile SME |  | x | x | Aravinda Manjappaiah | Application Developer – Advisory Services, Global Wealth & Asset Management Technology |

# Appendices

### Data dictionaries

### <https://mfc.sharepoint.com/:x:/r/sites/GWAMIDDLProjects/FAUM/Shared%20Documents/Analysis/BRD%27s%20Signed%20Off%20and%20Baselined/STM%27s%20Signed%20off%20and%20Baselined/Dataphile%20Column%20Metadata.xlsx?d=w9d35ce0917dd412591c46cd2190ddd46&csf=1&web=1&e=9DTMWU>

### DSA

See attached in sharepoint

### DSAdiscussion on trigger to start Dataphile ingestion

**From:** Aravinda Manjappaiah <[Aravinda\_Manjappaiah@manulife.ca](mailto:Aravinda_Manjappaiah@manulife.ca)>   
**Sent:** Tuesday, April 14, 2020 1:11 PM  
**To:** Vijay Sharma <[Vijay\_Sharma@manulife.ca](mailto:Vijay_Sharma@manulife.ca)>; John Steggerda <[John\_Steggerda@manulife.com](mailto:John_Steggerda@manulife.com)>  
**Cc:** Daniel Chang <[Daniel\_Chang@manulife.com](mailto:Daniel_Chang@manulife.com)>; Winnie Yu <[Winnie\_W\_Yu@manulife.com](mailto:Winnie_W_Yu@manulife.com)>; Naveen Kalathil <[Naveen\_Kalathil@jhancock.com](mailto:Naveen_Kalathil@jhancock.com)>  
**Subject:** RE: Dataphile Ingestion Schedule for IDDL

Hi John,

That is the right query to check the system update status.

If you run this right now, it would return the following value. If today’s date is returned, it means that the system date has been rolled over after the nightly processing.

**041420**

Thanks

**Aravinda Manjappaiah**

Application Developer – Advisory Services, Global Wealth & Asset Management Technology

**Manulife Investment Management**

**E** [Aravinda\_Manjappaiah@manulife.ca](mailto:Aravinda_Manjappaiah@manulife.com)

**T** 905-469-2100 x282116

**C** 416 938 8608

1235 North Service Road West,

Oakville, Ontario, Canada L6M 2W2



**From:** Vijay Sharma <[Vijay\_Sharma@manulife.ca](mailto:Vijay_Sharma@manulife.ca)>   
**Sent:** Tuesday, April 14, 2020 9:24 AM  
**To:** John Steggerda <[John\_Steggerda@manulife.com](mailto:John_Steggerda@manulife.com)>; Aravinda Manjappaiah <[Aravinda\_Manjappaiah@manulife.ca](mailto:Aravinda_Manjappaiah@manulife.ca)>  
**Cc:** Daniel Chang <[Daniel\_Chang@manulife.com](mailto:Daniel_Chang@manulife.com)>; Winnie Yu <[Winnie\_W\_Yu@manulife.com](mailto:Winnie_W_Yu@manulife.com)>; Naveen Kalathil <[Naveen\_Kalathil@jhancock.com](mailto:Naveen_Kalathil@jhancock.com)>  
**Subject:** RE: Dataphile Ingestion Schedule for IDDL

++ Naveen

**Thanks,**

**Vijay Sharma**  
Data Engineer  
**Manulife Investment Management  
  
E**    [vijay\_sharma@manulife.ca](mailto:vijay_sharma@manulife.ca)  
**T**    (647) 233  
  
250 Bloor Street East  
Torornto, Ontarion, Canada, M4W 1E6  


**From:** John Steggerda <[John\_Steggerda@manulife.com](mailto:John_Steggerda@manulife.com)>   
**Sent:** Monday, April 13, 2020 10:26 PM  
**To:** Aravinda Manjappaiah <[Aravinda\_Manjappaiah@manulife.ca](mailto:Aravinda_Manjappaiah@manulife.ca)>  
**Cc:** Vijay Sharma <[Vijay\_Sharma@manulife.ca](mailto:Vijay_Sharma@manulife.ca)>; Daniel Chang <[Daniel\_Chang@manulife.com](mailto:Daniel_Chang@manulife.com)>; Winnie Yu <[Winnie\_W\_Yu@manulife.com](mailto:Winnie_W_Yu@manulife.com)>  
**Subject:** FW: Dataphile Ingestion Schedule for IDDL

Aravinda

Can you elaborate on the query to tell if the system is ready?  I tried various combos of this query?

select code\_value from gncode where table\_name = 'Proc\_Dt'.

Maybe we need a quick call

cheers

John

### Data discussion points

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | Gap analysis done via meeting held with business and finance MVP rule applied since late in project. |  |
|  |  |  |
|  |  |  |
|  |  |  |