Auto Extraction and PIT collection via Documents [Going-Forward Basis]

(AUT001476)

May 2020

Project Owners- Rasneet Kaur/Anmole Dewan
Domain Leads – Paul Esposito/Srilekha Rathi

S&P Global

Agenda

- Background | Current Content Coverage and Maintenance
- Executive Summary – Going Forward Approach (Phase I)
- Russel 3K - Volume Analysis
- Russel 3K- Current Year's Financial Structures
- Auto Extraction Workflow

Background | Current Content Coverage and Maintenance

2 million Current Relationship are present in MI and **different teams** collect/update relationships

- ○ Relationships for only **~350K** children are actively maintained by the Hierarchy Management team annually.
- ○ ~**20%** of the total relationships are backed by transactions.

- Dual environment collection without proper Integrated Tools, Workflows and Pipelines.
- Both xCIQ and xSNL lack historical or point-in-time hierarchy data.

Current Hierarchy Coverage

- In additional to global, comprehensive corporate event coverage, including M&A, VC/PE, and Public Ownership, the following corporate structures are validated and maintained on a regular basis.
  - ○ Top to Bottom Hierarchy – S&P 1500 & S&P Global 100 (~250K children across ~10K Trees)
  - ○ New Security Issuers – Immediate Parent Info (~100 New Security Issuers/Day)
  - ○ Select Industry Coverage - (~100K Children across 55K Trees)

Point-In-Time Data for Relationships

- Only 5-7% of the entire universe has history of relationships (xSNL only)
- Start and End- Dates of relationships are not actively collected by the collection teams. 70% of the relationships do not have a start and end date associated with them.
- PIT Project would happen in phases over 2019 & 2020/2021 (depending on prioritization & resources)
  - NIC BankReg – *50K relationships under covered Banks (not all subsidiaries would be covered)
  - Transactions - *400K Transaction backed relationships
  - Russell 3K Documents – All subsidiaries disclosed by the Filer (*600K)

* There can be overlaps across the three sources – NIC, Transactions & Docs. Unique Quantum Study: TBD

Executive Summary | Auto Extraction and PIT – Going Forward Approach | Phase 1

1 Overview

1. Going – Forward Coverage Expansion from S&P 1500 Constituents to Russell 3K Constituents
2. Auto-compares documents 10K (Exhibit 21) with the database for last 2 years (current year with previous year)
3. Start and End Date collection

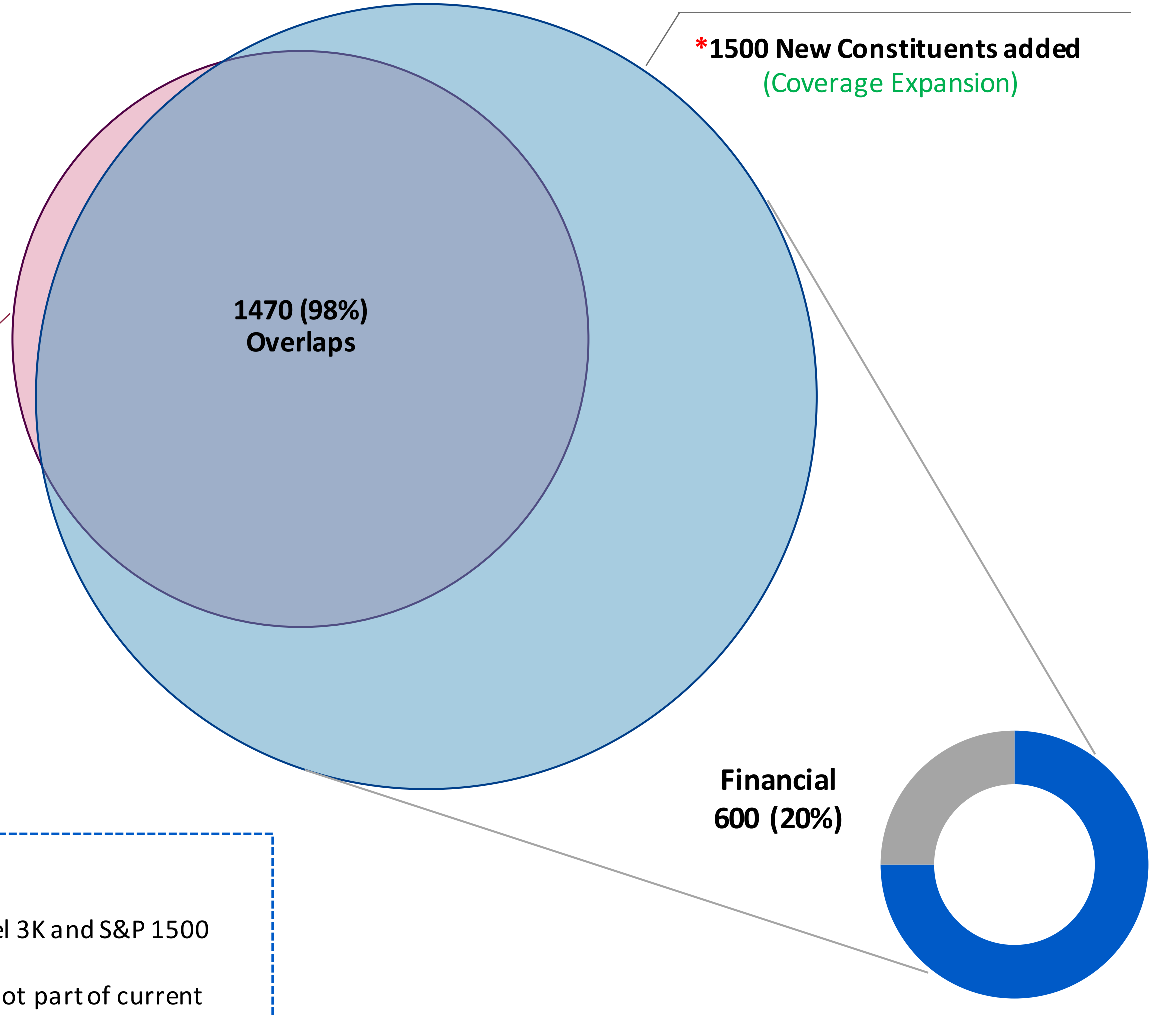
2 Scope

1. Subsidiaries filed in 10-K (Exhibit 21) belonging to Non-Financial Russel 3K constituents
2. Does not cover S&P Global 100 due to absence of Exhibit 21. Does not cover Financial Constituents as NIC/NAIC are more comprehensive sources for the same.
3. **Phase I Scope** – Non Financial Russell 3K Constituents where the structure type is- 'Flat Structure without Stake information' (clean structures). Details here.
4. Quantum in Phase I
   a. Structures – 1200- 1300 corporate trees
   b. Subsidiaries – Work in Progress

3 Value Proposition

1. Coverage Expansion to Russell 3K Constituents – End to End Workflow Solution
2. Point-In-Time Data - Date on which parent first invested in the child (Start date). Start Date would primarily be the Period Date/Filing Date and Date Source " Document Date"
3. Faster and more efficient way of performing repeated tasks with reduced manual intervention (Automated to Manual Ratio: TBD)
4. Dashboard that collates stats from the model such as – Filings received/subsidiaries ingested/manual review needed, etc.
5. Source Tagging – Tagging a source from which a relationship has been added/updated. [R&D in Progress]

Russell 3K Constituents | Volume Analysis



*1500 New Constituents added
(Coverage Expansion)

30 (2%)
of S&P 1500 is not a part of Russell 3K

1470 (98%)
Overlaps

Financial
600 (20%)

Non-Financial
2430 (80%)

🟦 Russel 3K Constituents

🟥 S&P 400,500 and 600 constituents

*Investment Firms (except REITs) would be excluded from the scope -*90 constituents = ~2950 structures*

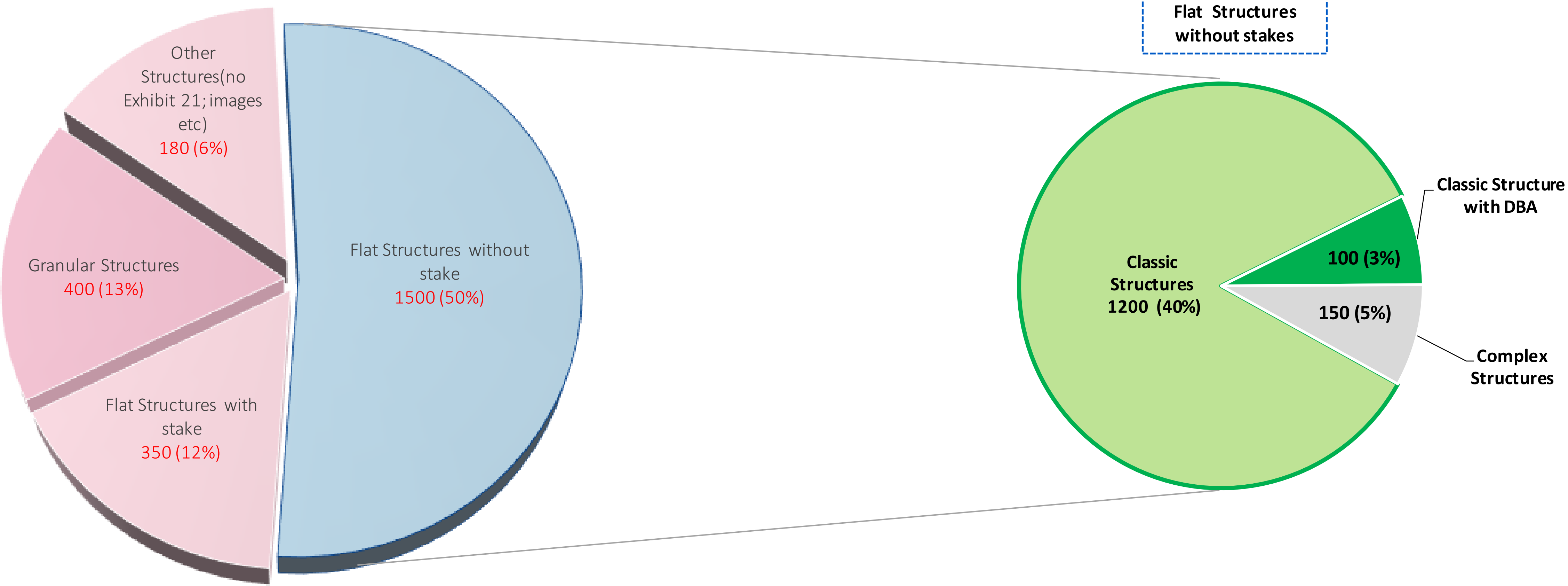
Going Forward in PIT approach – Phase 1:

1. **Coverage Expansion overall scope**: Complete Russel 3K and S&P 1500 (excluding non REIT Investment Firms)
2. **Financial structures require NIC and NAIC Filings:** Not part of current release.

Russel 3K Constituent | Current Year's Non-Financial structures

2430 (80%) *
Non-Financial Constituents

1450 (48%) **
Flat Structures without stakes



*The base of percentage is the complete Russel 3K (3000 companies)

** When both years are taken together, there are 1450 Flat without stake structures. (3% excluded)

Scope - Going Forward in PIT approach – Phase 1:

1. Non Financial structures
2. Flat without stakes - ~1300 (43%) [**S&P 1500** = 800 | **Unique Russell 3K** = 500 (tentatively)]
   a. Classic structures ~ 1200 structures
   b. Classic structures with DBA ~ 100 structures
3. Complex structures not covered in current release

Workflow | Auto Extraction and PIT – Going Forward Approach | Phase 1

Step 1 New Filing Alert

New Filing Trigger

Accessing **Exhibit 21 in HTML** format

2019 & 2020

Step 2 Auto - Extraction

2019 = 2020

Auto-extracted output format

1. **Structure Identification**
2. **Auto Extraction** of subsidiaries from last 2 years' Exhibit 21
3. **Data Cleansing**
4. **Keyword Flagging**

Child Name	Jurisdiction	Stakes	Parent Company Name	Ultimate Parent Name	Year	Filing Date	As of or Period Date
Microsoft India Pvt Ltd.	India	NA	Microsoft, Inc.	Microsoft, Inc.	2020	3-Mar-2020	31-Dec-2019
Microsoft India Pvt Ltd.	India	NA	Microsoft, Inc.	Microsoft, Inc.	2019	2-Mar-2019	31-Dec-2018

Step 3 Company Mapping

KENSHO

Entity Identification & Mapping

1. Add-News
2. Add-Existing
3. Manual Review via **Pathfinder Jobs**
4. Automated vs Manual Ratio

Step 4 Doc to Doc Comparison

2019 v/s 2020

1. **Identification of Overlaps/Additions/Subtractions** of subsidiaries from 2019 and 2020
2. **Creation of PIT enabled relationship data.**

Step 5 Doc Output to DB Comparison

2019/2020 Document v/s Database: Relationships data + Transaction data

Step 6 Data Ingestion/Updates

1. Queuing doubtful cases for manual review via Pathfinder.
2. Auto Insert/Updates into DMS/CIQ using GB Tool or BOT

Key Highlights:

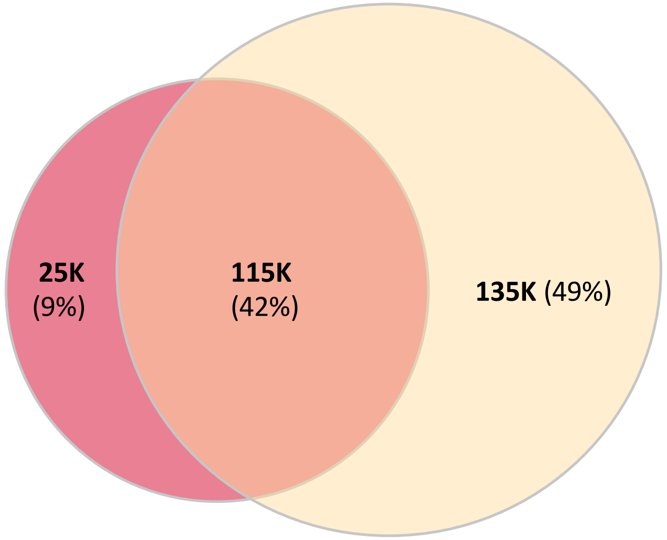
- ✓ **End to end** workflow solution.
- ✓ Processing failure will raise **Pathfinder job** for that particular step to be reviewed manually. Rest of the steps will remain automated.
- ✓ **Real time Dashboard** will display daily performance of the model along with stats.
- ✓ **Source Tagging** of relationships.

Resource Requirement

Coverage Analysis

Coverage/Quantum Analysis | Russell 3K Constituents

Overlapping with S&P 1500

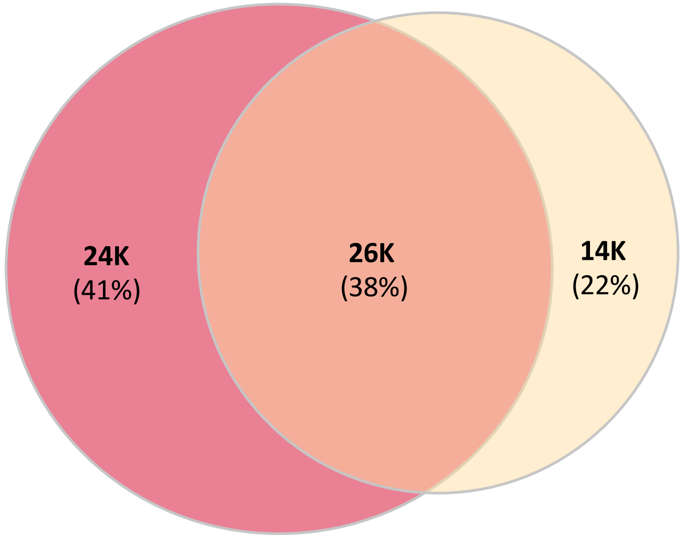


Key Takeaways:

1. Number of Filers = 1500 (Financial + Non-Financial)
2. Average tree size basis Current year's Exhibit 21* = ~90
3. Total Subsidiaries in Exhibit 21 = 140K (Number of Filers X Average tree size)
4. Total Subsidiaries in Database = 250K (Actual counts from CIQ Database)
5. Total Subsidiaries under Filer ((Document + Database) – *Overlaps) = 275K (140K + 250K – 115K)
6. Net New Companies (Only Document)* = **25K**. This includes only Additions & Overlaps from previous year document. Removals have not been considered.

*Sourced from Rev 10
Does not include NIC/NAIC Counts/Website, etc.

Unique Russell 3K (Not Overlapping with S&P 1500) - New



Key Takeaways:

1. Number of Filers = 1500 (Financial + Non-Financial)
2. Average tree size basis Current year's Exhibit 21 = 35 (700 structures sampled)
3. Total Subsidiaries in Exhibit 21 = 50K (Number of Filers X Average tree size)
4. Total Subsidiaries in Database = 40K (Actual counts from CIQ Database)
   - Merged Entities = 7K & Current Subsidiary/Current Investment Arm = 33K
   - Asset Products and Funds have been excluded.
5. Total Subsidiaries under Filer ((Exhibit 21 + Database) – **Overlaps) = ~65K (50K + 40K – 25K)
6. Net New Companies (Exhibit 21)* = **24K**. This includes only Additions & Overlaps from previous year document. Removals have not been considered.

*Net New Companies % is a guesstimate basis 30 structures sampled
**Overlap % is a guesstimate basis 30 structures sampled

S&P Global

●
- Only Document
- Net New Companies*

○ Only Database

● Overlaps in Document & Database

Manual vs Automated Universe

Unique Russell 3K Constituents (Non Financials) – Non Covered* | Manual Review & FTE Requirement

*1200 Structures, ~55K subsidiaries

Category 1

Auto Extraction Failure/Footers

Manual Review Scope = 360 filers (30% filers)

Category 2

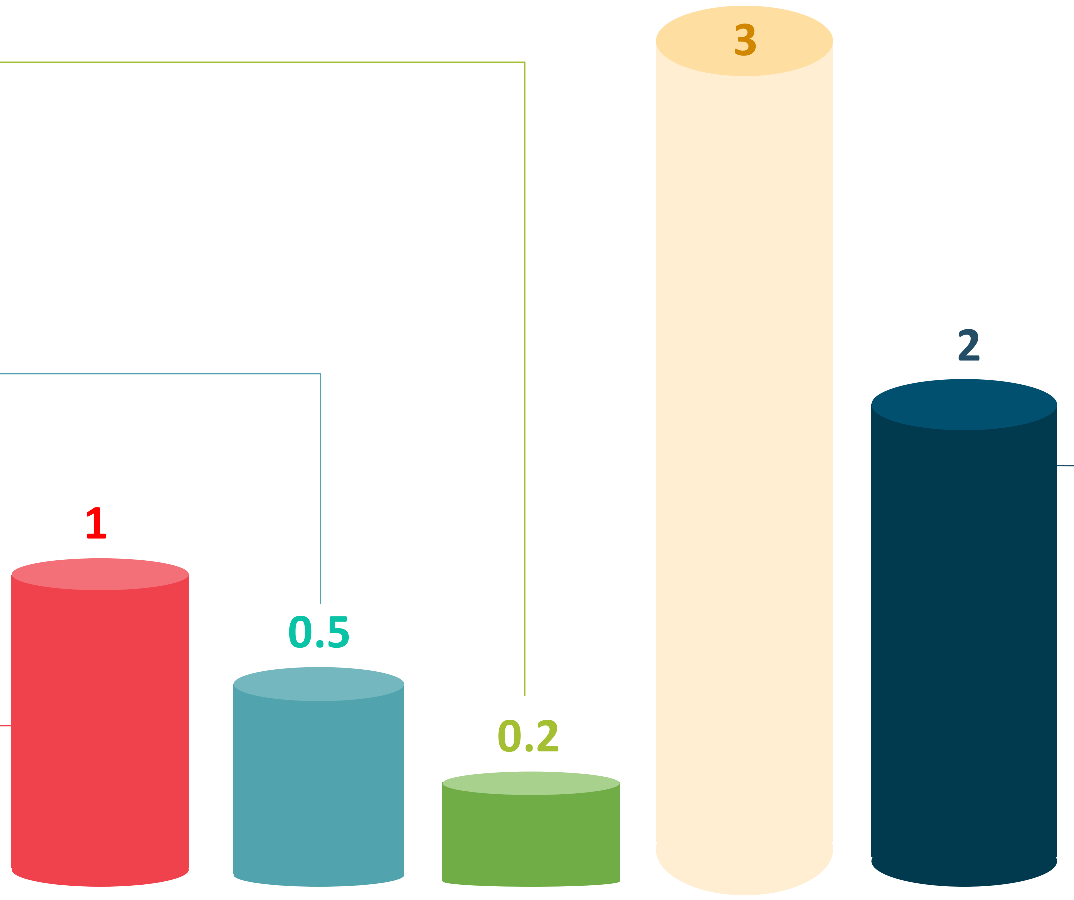
KENSHO - Company Mapping (Middle Buckets)

Manual Review Scope = 9000 subsidiaries (18%)

Category 3

Discrepancies between Exhibit 21 and Database

Manual Review Scope = 5K subsidiaries (25%)



Total FTE Requirement = 7

Both Doc & DB

Category 5

Client Questions/ Training/QA

Category 4

Only database & Celsus Search

Manual Review Scope = 18K subsidiaries

S&P Global

Unique Russell 3K Constituents – Phase 1* | Manual Review & FTE Requirement

*Flat structures without stakes + Non Financial/Non REIT = 500 structures, ~25K Subsidiaries

Category 1

Auto Extraction Failure/Footers

Manual Review Scope = 50 filers (10% filers)

Category 2

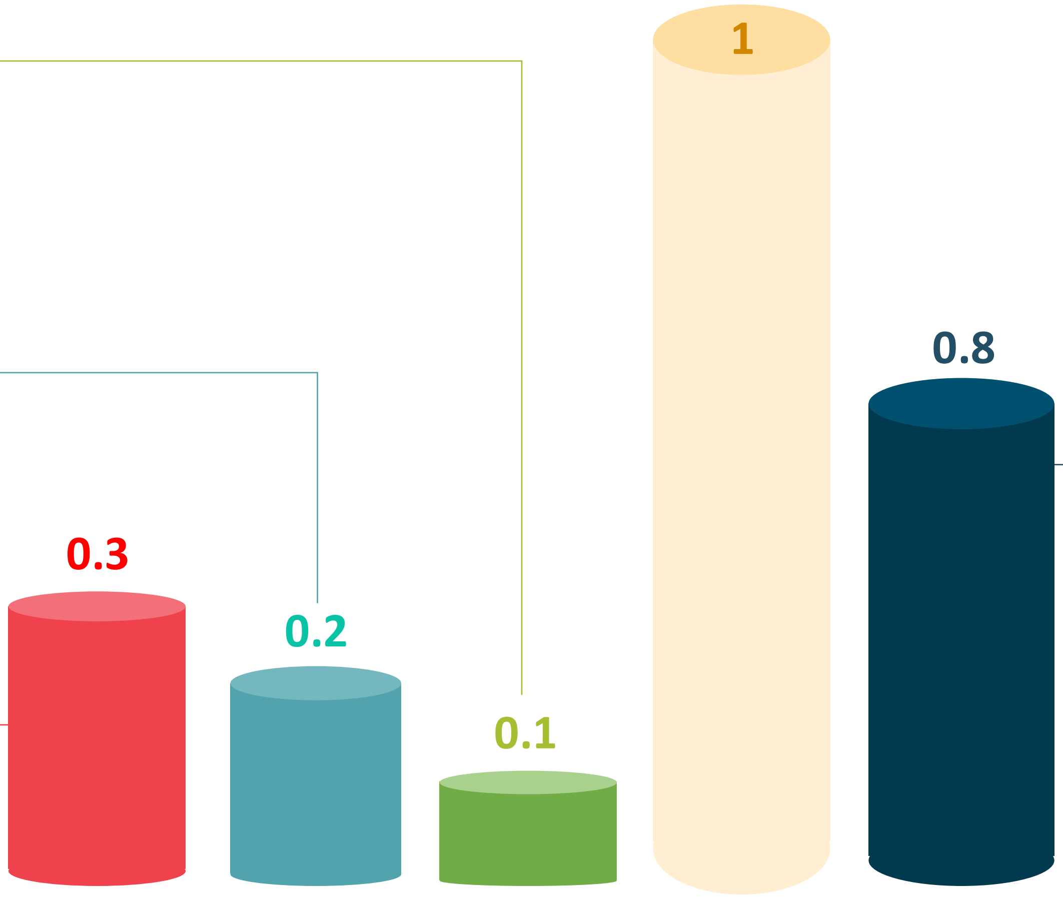
KENSHO - Company Mapping (Middle Buckets)

Manual Review Scope = 3000 subsidiaries (18%)

Category 3

Discrepancies between Exhibit 21 and Database

Manual Review Scope = 2K subsidiaries (25%)



Category 5

Client Questions/Training/QA

Category 4

Only database & Celsus Search

Manual Review Scope = 6K subsidiaries

Total FTE Requirement = 3

Unique Constituents | Manual Review

Auto Extraction: Overall scope of manual review is **~16%**.

1. Keyword "Exhibit 21" not present in the document. Quantum = 17 filers (~1%)
2. Multicolumn subsidiary list. Quantum = 5 filers (0.3% of 1500)
3. Document table structure not in required format. Quantum = 25 filers (~2% of 1500)
4. filers (0.5% of 1500)
5. Structures with relevant footnotes. Quantum = 34Keyword "Exhibit 21" present on multiple pages. 8 (2.3% of 1500)
6. Structure type Flat with Jurisdiction but, jurisdiction missing for some entities. TBD
7. 10K filing of the previous year not available. Quantum = ~90 filers (6% of 1500)

KENSHO - Company Mapping: Overall scope of manual review is **~18%**.

1. Subsidiaries with KENSHO confidence results >20 and <90. Quantum = 8K (15% of 50K)
2. Subsidiaries with KENSHO confidence results >90. Quantum = 1500 (3% of 50K)

Document Vs Database: Overall scope for manual review comes out to be ~**25%**. This covers the following scenarios:

1. Missing Stakes. Quantum = 80 subsidiaries (0.4%)
2. Ultimate Parent Mismatch. Quantum = 600 subsidiaries (3%)
3. Direct Parent Mismatch. Quantum = 4400 subsidiaries (~22%)
4. Relationship/Stake type mismatch. Quantum = 80 subsidiaries (0.4%)

Product Questions

Open Questions | Product

1

Exhibit 21 as the Base (Primary) Source

1. **For new constituents (Expansion)** -
   - As of now, only look at Exhibit 21 as the main source for Non-Financial Industries. Subsequently, look at automating other sources as well such as 8Ks, 10Q, etc.
   - For Financial Structures – Automate NIC/NAIC along with Exhibit 21

2. **For S&P 1500 (Existing Workflow)**
   - No change in the workflow in terms of sourcing. Continue with the robust collection of looking at all sources such as Website, Press Releases, 8Ks, etc.
   - Start the collection of start and end dates

2

Do we want the expansion on both CIQ and MI?

- Currently, there is a Forward Data Pipeline for Relationships but, no Reverse Data Pipeline.
- Our suggestion is to expand on CIQ and then flow the data to MI via Forward Data Pipeline.
- Clarity on this aspect will help us understand and make decisions regarding the Ingestion and Collection strategy.
- In Phase I, idea is to look at expansion along with collection of start and end dates. Evolution of shareholding/history would not be captured in Phase I.

3

What SLAs are we looking at for the Russell 3K Expansion (new constituents)

- For S&P 1500, the SLA varies depending on the market capitalization and the peak period.
- Generally between 10-20 days during Q1 and it reduces subsequently.
- Questions –
  1. What timeliness guarantee are we looking at for the new expansion?
  2. Are we also looking at guaranteeing start and end dates to our clients for 2021?

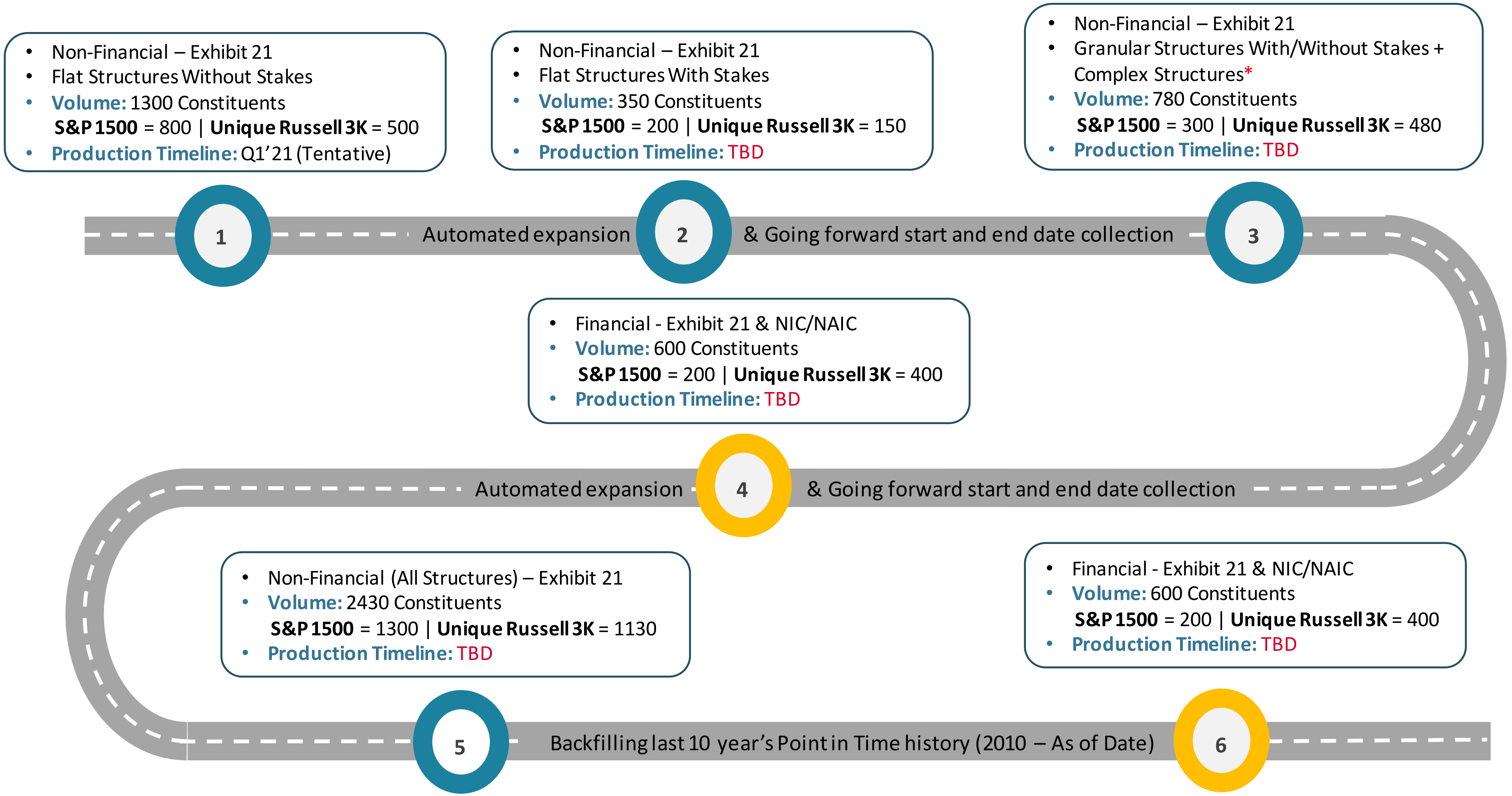
4

Daily Refresh for Russell 3K Constituents

Following are some points on constituent changes for Russell 3K:

- Idea is to go for a daily alert mechanism such that we are notified if there is any -
  a. Addition to the Russell 3K Constituent List, or
  b. If there is a new 10K for an existing Russell 3K Constituent
- Once a Filer is no longer a constituent of Russell 3K, we would not maintain the corporate structure of the same.

Project Roadmap



- Non-Financial – Exhibit 21
- Flat Structures Without Stakes
- **Volume:** 1300 Constituents
  S&P 1500 = 800 | **Unique Russell 3K** = 500
- **Production Timeline:** Q1’21 (Tentative)

- Non-Financial – Exhibit 21
- Flat Structures With Stakes
- **Volume:** 350 Constituents
  S&P 1500 = 200 | **Unique Russell 3K** = 150
- **Production Timeline:** TBD

- Non-Financial – Exhibit 21
- Granular Structures With/Without Stakes + Complex Structures*
- **Volume:** 780 Constituents
  S&P 1500 = 300 | **Unique Russell 3K** = 480
- **Production Timeline:** TBD

1 Automated expansion 2 & Going forward start and end date collection 3

- Financial - Exhibit 21 & NIC/NAIC
- **Volume:** 600 Constituents
  S&P 1500 = 200 | **Unique Russell 3K** = 400
- **Production Timeline:** TBD

Automated expansion 4 & Going forward start and end date collection

- Non-Financial (All Structures) – Exhibit 21
- **Volume:** 2430 Constituents
  S&P 1500 = 1300 | **Unique Russell 3K** = 1130
- **Production Timeline:** TBD

- Financial - Exhibit 21 & NIC/NAIC
- **Volume:** 600 Constituents
  S&P 1500 = 200 | **Unique Russell 3K** = 400
- **Production Timeline:** TBD

5 Backfilling last 10 year’s Point in Time history (2010 – As of Date) 6

Key Points to note:

- Till Phase 4, extraction of 15% of the structures i.e. ~450 structures can’t be automated.
- The new expansion is focused towards Exhibit 21 whereas for S&P 1500, all sources are being covered.
- Therefore, the effort for both these indices in each Phase is different as we have to develop two models.

Russell 3K
ansion - Body of W

* Different type of structures across two years. Complicated footers

Financial Vs Non-Financial | Unique Russell 3K Constituents

		Flat	%	Indented	%	Exhibit 21 Not Available	%	Total Unique Russell 3K Structures
1	Financial Structures	185	60%	120	39%	5	1%	310
1A	Banks	110	60%	70	39%	3	1%	183
1B	Insurance	20	60%	13	37%	1	3%	34
1C	Others	55	60%	37	39%	1	1%	93
2	REITs	75	94%	0	0%	5	6%	80
3	Non-Financial Structures	730	66%	80	7%	300	27%	1110
	Total Unique Russell 3K Structures	**990**	**66%**	**200**	**13%**	**310**	**21%**	**1500**

- ✓ For Category 1A, NIC will have to be considered too.
- ✓ For Category 1B, NAIC will have to be considered too.
- ✓ The Auto Extraction of ~730 Non-Financial flat structures as a part of Phase 1 is complete.

Unique Russell 3K Constituents | Data Point Level Accuracy

Accuracy numbers given below are based on a **sample of 215 structures** which were tested.

Table 1: Auto Extraction

S. No.	Data Point	Missing %	Accuracy %
1	Child Name	2%	99.7%
2	Jurisdiction	1%	100%
3	Direct Parent Name	-	99.5%
4	Additional Information	0.5%	99.9%
5	DBA/Historical Name	-	99.7%
6	Stakes	-	100%
7	Date of Filing	-	100%
8	As of or Period Date	-	100%

Table 2: KENSHO

KENSHO Bucket	Accuracy %
Add New	98%
Add Existing	97%

Key Pointers – Table 1:

1. **Data point 1:** ~2% of missing child/subsidiary names is to account for structures where the keyword "Exhibit 21" in mentioned on multiples pages. Example: https://www.sec.gov/Archives/edgar/data/1286225/000128622520000011/exhibit211201910-k.htm

2. **Data point 2:** In ~1% of cases, jurisdiction information given in the Exhibit 21 might not be extracted by the model. Example, our model recognizes Country = Bolivarian Republic of Venezuela but, in Exhibit 21 Country = Venezuela, Bolivarian Republic of.

3. **Data point 3:** ~0.5% of Direct Parent Name can be expected to be incorrectly identified. This will happen when a Granular structure is recognized as a Flat Structure. Wrong structure type identification can only happen in such scenarios where list of subsidiaries is given without any indentation but, headings give granular information. Example: https://www.sec.gov/Archives/edgar/data/1157408/000155837019007293/lrn-20190630ex211f86d37.htm

4. **Data point 4:** Additional information refers information related to status/granular parent/type of subsidiary – acquisition corp., holding company, given in the exhibit 21. This information can be missed if the keyword given is not identified by the model.

Project Progress and Volumes

✓ **Total Structures in Phase 1** = 680 (Unique Russell 3K, Non-Financial, Non-REIT, Flat Without Stakes Structures)

✓ **Total Subsidiaries** = 24K (Part of current year Exhibit 21)

✓ **Estimated Production Release** = January 2021

✓ **Automated Vs Manual Review Scope** = 60% Automated & 40% Manual Review.

✓ If manual resources are not available, we'll be able to complete only 50% of the structures.

✓ **Total Structures in Phase 2** = 70 (Unique Russell 3K, Non-Financial, Non-REIT, Flat With Stakes Structures)

✓ **Total Subsidiaries** = 2.5K (Part of current year Exhibit 21)

✓ We are pushing to achieve production release of Phase 1 & 2 together by January 2021

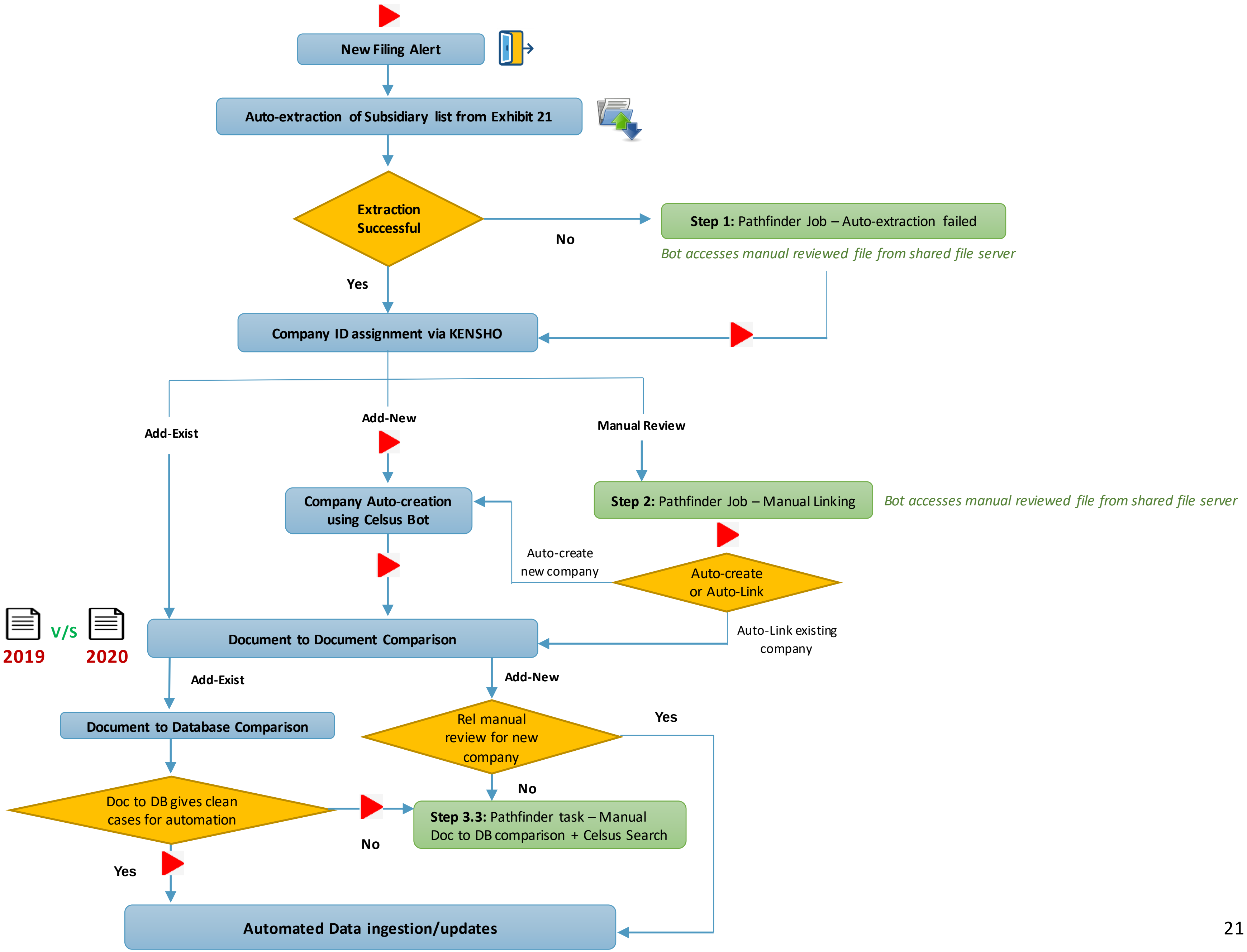
✓ **Savings in terms of FTEs**:

FTE Requirement	
After Automation	2.20
Pure Manual Review	4.95
Savings	2.75
	55.59%

Pathfinder Workflow | Unique Russell 3K Non-Financial Constituents

Document = Current year Exhibit 21

Database = CIQ Relationships



New Filing Alert

Auto-extraction of Subsidiary list from Exhibit 21

Extraction Successful

No

Step 1: Pathfinder Job – Auto-extraction failed

Bot accesses manual reviewed file from shared file server

Yes

Company ID assignment via KENSHO

Add-Exist

Add-New

Manual Review

Company Auto-creation using Celsus Bot

Step 2: Pathfinder Job – Manual Linking

Bot accesses manual reviewed file from shared file server

Auto-create new company

Auto-create or Auto-Link

Auto-Link existing company

2019 v/s 2020

Document to Document Comparison

Add-Exist

Add-New

Document to Database Comparison

Rel manual review for new company

Yes

No

Doc to DB gives clean cases for automation

No

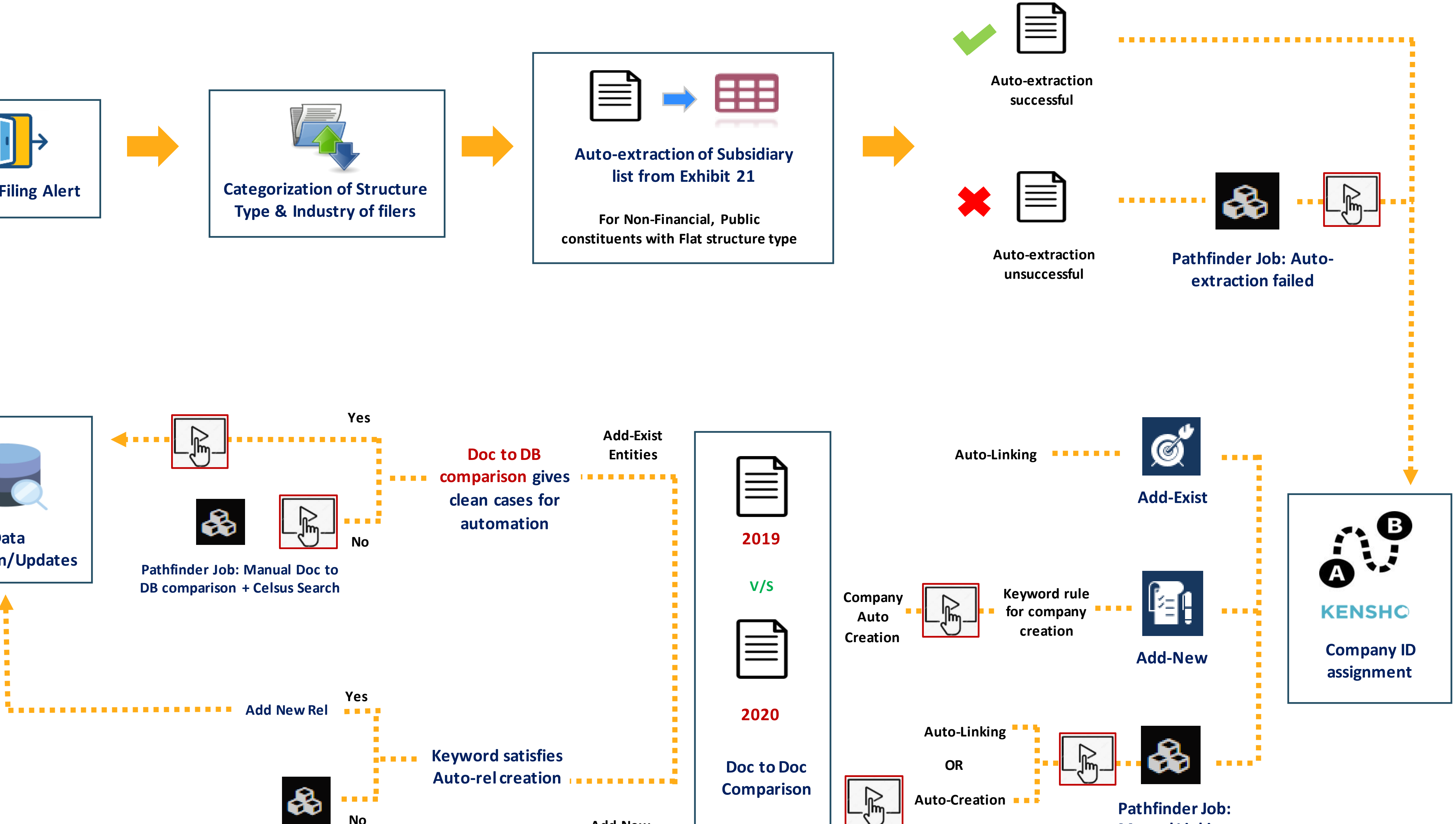
Step 3.3: Pathfinder task – Manual Doc to DB comparison + Celsus Search

Yes

Automated Data ingestion/updates

Annexure

der Workflow | Unique Russell 3K Non Financial Constituents



Filing Alert

Categorization of Structure Type & Industry of filers

Auto-extraction of Subsidiary list from Exhibit 21

For Non-Financial, Public constituents with Flat structure type

Auto-extraction successful

Auto-extraction unsuccessful

Pathfinder Job: Auto-extraction failed

KENSHO

Company ID assignment

Auto-Linking

Add-Exist

Company Auto Creation

Keyword rule for company creation

Add-New

Auto-Linking

OR

Auto-Creation

Pathfinder Job:

2019

V/S

2020

Doc to Doc Comparison

Add-Exist Entities

Doc to DB comparison gives clean cases for automation

Yes

No

Pathfinder Job: Manual Doc to DB comparison + Celsus Search

ata

n/Updates

Keyword satisfies Auto-rel creation

Yes

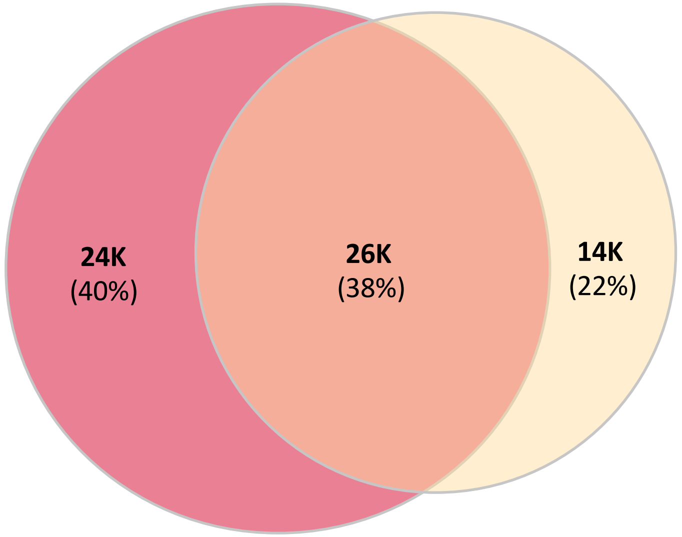
Add New Rel

No

Add New

nent VS Database Coverage Analysis | Unique Russell 3K Constituents

Unique Russell 3K (Not Overlapping with S&P 1500)



Key Takeaways:

1. Number of Filers = 1500 (Financial + Non-Financial)

2. Average tree size basis Current year's Exhibit 21 = 35 (700 structures sampled)

3. Total Subsidiaries in Exhibit 21 = 50K (Number of Filers X Average tree size)

4. Total Subsidiaries in Database = 40K (Actual counts from CIQ Database)
   - Merged Entities = 7K & Current Subsidiary/Current Investment Arm = 33K
   - Asset Products and Funds have been excluded.

5. Total Subsidiaries under Filer ((Exhibit 21 + Database) – **Overlaps) = ~65K (50K + 40K – 25K)

6. Net New Companies (Exhibit 21)* = **24K**. This includes only Additions & Overlaps from previous year document. Removals have not been considered.

*Net New Companies % is a guesstimate basis 30 structures sampled
**Overlap % is a guesstimate basis 30 structures sampled

● Only Document ○ Only Database ● Overlaps in Document & Database

nent Vs Database | Unique Constituents – Net New Companies

Net New Companies (From Exhibit 21)	24K
Non-Financial	19K
Financial	5K

Accuracy Analysis	
Category	Count of Subsidiaries Sampled
Only Document	**315**
Net New Companies	315

315 Net New Companies from Exhibit 21 across **all Non financial structures types** were studied.

Key Takeaways:

On Outside Research:

1. Granular parent information is available for **2%** (8 entities) of 315 Net New Companies. Sources providing this information were 10 K, 10 Q and Form DEF 14A.

2. Out of the total analyzed structures, stakes information is available for **8%** (27/315 entities).

Open Question:

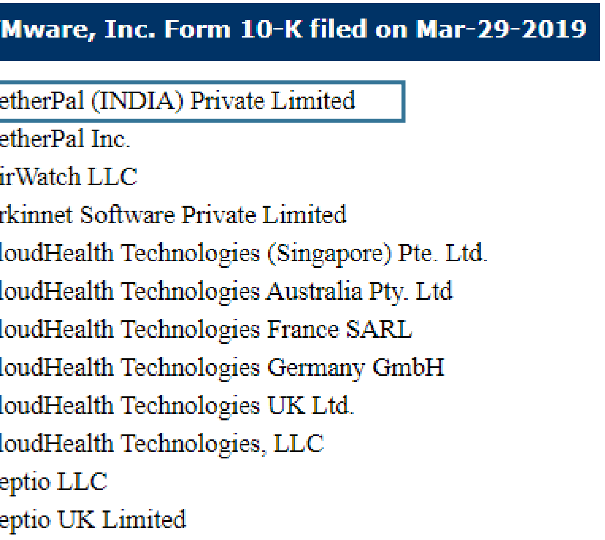
Option 1: All Net New Companies are auto ingested as reported in the exhibit
Resource Requirement = 0 FTEs. Risk on data quality = Stakes info missed out for ~1500/19K entities and Granular parent info missed out for ~400/19K entities.

Option 2: Manually review all Net New Companies
Resource Requirement = 4 FTEs with a risk of 4-5% on data quality due to human mistakes.
If going with manual review of these entities, what should be the workflow?

1. All net new companies are manually reviewed at the filer level
2. At first, the relationship is aligned according to exhibit and then, these entities are manually

ew Companies | Start Date & Data Source

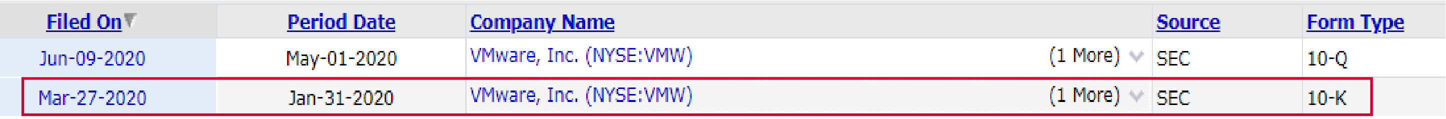
2019 Exhibit 21



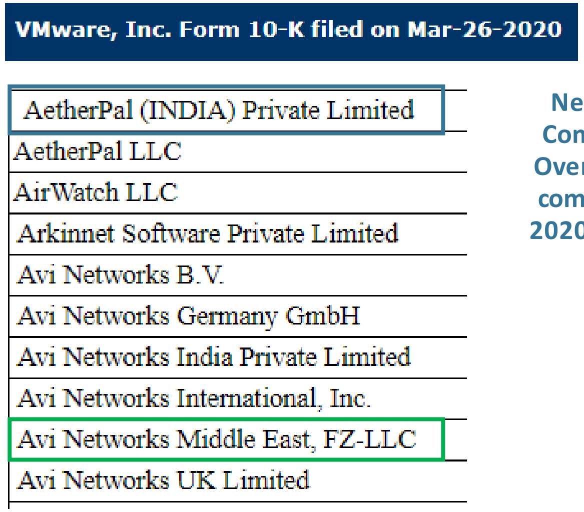
Mware, Inc. Form 10-K filed on Mar-29-2019

etherPal (INDIA) Private Limited
etherPal Inc.
irWatch LLC
rkinnet Software Private Limited
loudHealth Technologies (Singapore) Pte. Ltd.
loudHealth Technologies Australia Pty. Ltd
loudHealth Technologies France SARL
loudHealth Technologies Germany GmbH
loudHealth Technologies UK Ltd.
loudHealth Technologies, LLC
eptio LLC
eptio UK Limited

2020 Exhibit 21



Filed On	Period Date	Company Name		Source	Form Type
Jun-09-2020	May-01-2020	VMware, Inc. (NYSE:VMW)	(1 More)	SEC	10-Q
Mar-27-2020	Jan-31-2020	VMware, Inc. (NYSE:VMW)	(1 More)	SEC	10-K



VMware, Inc. Form 10-K filed on Mar-26-2020

AetherPal (INDIA) Private Limited
AetherPal LLC
AirWatch LLC
Arkinnet Software Private Limited
Avi Networks B.V.
Avi Networks Germany GmbH
Avi Networks India Private Limited
Avi Networks International, Inc.
Avi Networks Middle East, FZ-LLC
Avi Networks UK Limited

Net New Company - Overlapping company in 2020 & 2019

Net New Company, Not present in 2019

I. Start Date = 1/1/2015 and Data Source = Unknown Date
II. Start Date = 1/31/2020 and Data Source = Document Date
III. **80 - 85%** of the Net New Companies are a part of both Exhibit 21 of 2020 & 2019. For these ~250 companies, start date will go as 1/1/2015 with Date source as Unknown date. For the rest, start date will go as 2020 Document Date/Period date and Date source will be Document date.
IV. **Extrapolated figures:** Start date = 1/1/2015 for ~20K entities with Date source as Unknown date. For ~4K, Start date = Filing date/Period date with Date source as Document date.
V. Once last 10 year's historical documents are covered, then the start date of these 80% entities would get changed from 1/1/2015 to Document date.
VI. **Ingestion Strategy:** These Net New Companies& their associated relationships should be ingested in CIQ and they will flow to MI via Forward Data Pipeline

   a. This would ensure data availability on both CIQ and MI platform
   b. Accessibility on Xpressfeed.

istorical/Alternate Names Analysis | Russell 3K Constituents

Open Question:

1. For Exhibit 21s that give us Subsidiaries with DBA/Historical/AKA names, For Add-New Cases, is there a need to manually review them to avoid creating duplicates.

Coverage Quantum:

1. Total Number of Filers = 3000
2. Total No of Structures with DBA information = 200 (6%) (based on sampling given below)
   a) Total Unique Constituents with DBA information = 80 (5%) (based on 600 structures sampled)
   b) Total S&P 1500 Constituents with DBA information = 120 (8%) (based on 1000 structures sampled)

Analysis Methodology:

1. This Analysis is based on sampling of 138 subsidiaries with DBA/Alternate names/Historical Names information.

2. A total of 4 Unique and 3 S&P 1500 constituents were studied for the analysis

Key Takeaways:

Kensho Confidence Range	Errors Caused Due to Absence of DBA names in Database for Add-New	Extrapolated subsidiaries which can lead to dupes based on DBA information
0-10	2	110
Total Subsidiaries	**25**	**3000* out of 20,000**
Percentage	8%	
10-20	2	160
Total Subsidiaries	15	1500* out of 10,000
Percentage	13%	
Total (20-100)	98	-
Total Subsidiaries	138	-

1. For Unique Russel 3K constituents, the duplications found due to Subsidiaries having Alternate names in the Exhibit 21 were not found.

2. For S&P 1500
   a) Subsidiaries which returned 0-10 Kensho confidence exists in Database with alternate names only ~ 8% (110 out of 3000)
   b) Subsidiaries which returned 10-20 Kensho confidence exists in Database with alternate names only.~13% (160 out of 1500)

istorical/Alternate Names Analysis | Russell 3K Constituents

Proposal

Option 1: All Net New Companies with unique DBA information are auto ingested as reported in the exhibit**
Resource Requirement = 0 FTEs. Risk on dupes = 4% to 13% with subsidiaries with DBA information

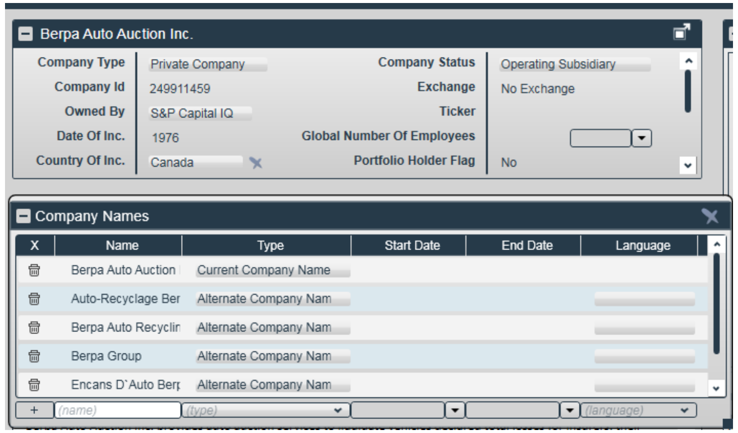
Option 2: Manually review all Net New Companies with unique DBA information only for S&P 1500
Resource Requirement = 0.35 FTE (75 days total for 1 FTE) with a risk of 2-3% on data quality due to human mistakes.

**What is Unique alternate name- If a subsidiary has an alternate name which is not given to another subsidiary in the same Filer

Errors caused by absence of DBA:

1. Information missed by Researcher (Status & Exist updated)
2. DBA information not added in alternate names
3. Company not searched using DBA names to link the subsidiaries name with their DBA

Name of subsidiary	State of incorporation	Name(s) under which subsidiary does business	UP	CIQID using names and country	CIQID using DBA	Kensho ciqid	Kensho confidence	Subsidiary Name (DB)
Copart Montréal Inc.	Canada	Copart Auction, Berpa Auto Auction, GPS Secure Storage, Encan Copart, Encan D'Autos Berpa, GPS Entreposage Sécuritaire, Réseau Des Commerçants Automobiles Accrédités Du Québec	Copart, Inc. (NasdaqGS:CPRT)		249911459	216344085	7.58	Coparts

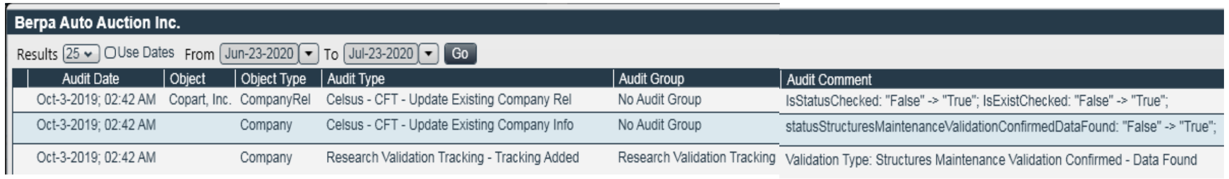


Berpa Auto Auction Inc.

Company Type	Private Company	Company Status	Operating Subsidiary
Company Id	249911459	Exchange	No Exchange
Owned By	S&P Capital IQ	Ticker	
Date Of Inc.	1976	Global Number Of Employees	
Country Of Inc.	Canada	Portfolio Holder Flag	No

Company Names

X	Name	Type	Start Date	End Date	Language
	Berpa Auto Auction	Current Company Name			
	Auto-Recyclage Ber	Alternate Company Nam			
	Berpa Auto Recyclin	Alternate Company Nam			
	Berpa Group	Alternate Company Nam			
	Encans D`Auto Berp	Alternate Company Nam			



Berpa Auto Auction Inc.

Results 25 ○Use Dates From Jun-23-2020 To Jul-23-2020 Go

Audit Date	Object	Object Type	Audit Type	Audit Group	Audit Comment
Oct-3-2019; 02:42 AM	Copart, Inc.	CompanyRel	Celsus - CFT - Update Existing Company Rel	No Audit Group	IsStatusChecked: "False" -> "True"; IsExistChecked: "False" -> "True";
Oct-3-2019; 02:42 AM		Company	Celsus - CFT - Update Existing Company Info	No Audit Group	statusStructuresMaintenanceValidationConfirmedDataFound: "False" -> "True";
Oct-3-2019; 02:42 AM		Company	Research Validation Tracking - Tracking Added	Research Validation Tracking	Validation Type: Structures Maintenance Validation Confirmed - Data Found