

Electronic Searches at CDC Discussion with CDC-FOIA

Srinath Tutukuri

IT Project Manager, FOIA, CDC

OGIS/CDC FOIA Discussion

05/06/21



Agenda

FOIA Enterprise Searches(ES)

- ES Overview
- ES Categorization
- ES Issues
- Improve ES
- Q & A

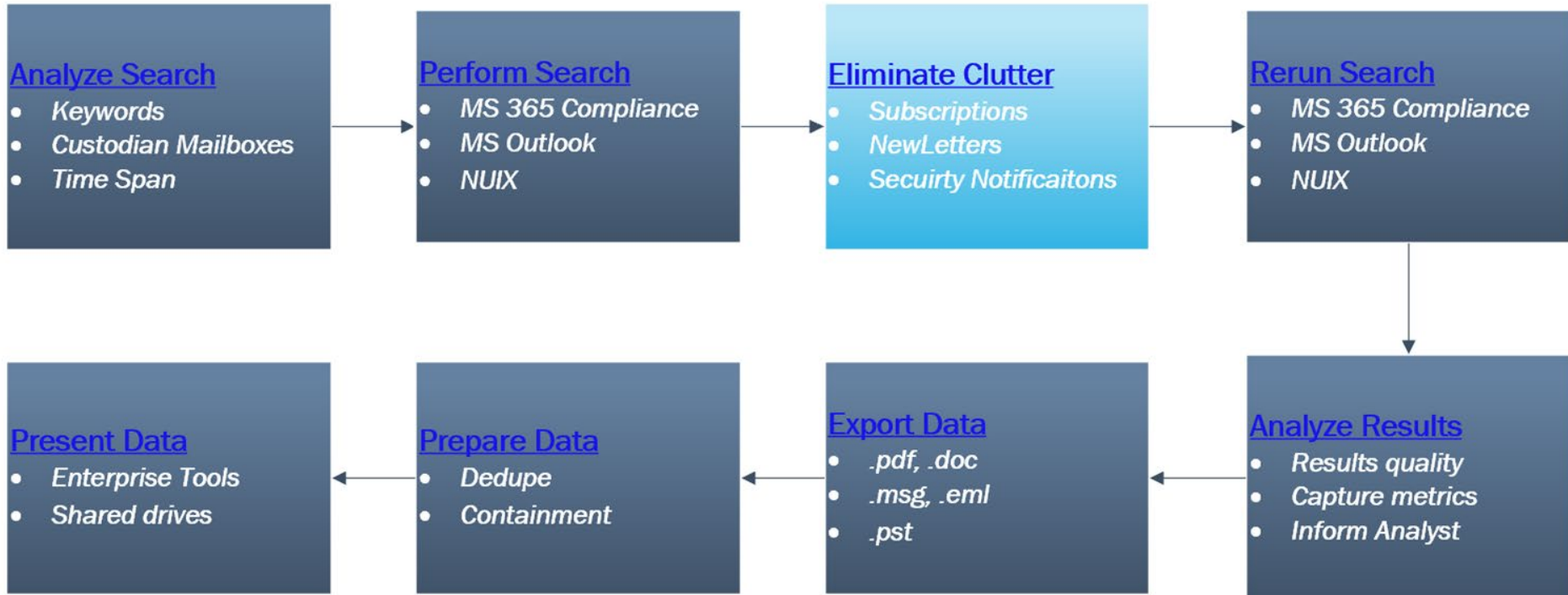


ES Overview

- ES Process Flow
- ES Tools
- ES Dedupe & Containment



ES Process Flow

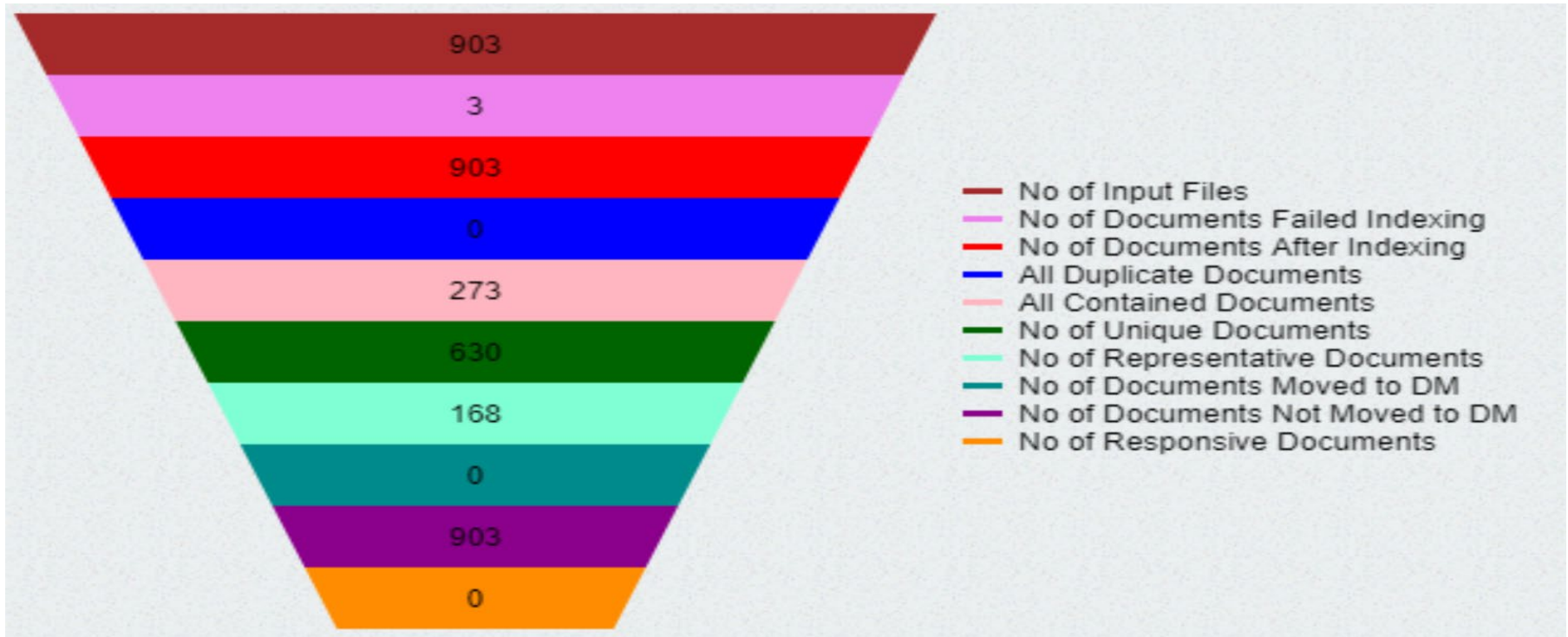


ES Tools

- Microsoft 365 Compliance
- Microsoft Outlook
- FOIA Express
- NUIX



ES Dedupe & Containment



Categorize Enterprise Search

- Low Intensity
- Moderate Intensity
- High Intensity



Low Intensity Search

- Custodian mailboxes defined.
- Short date range.
- Limited Participants.
- Unambiguous Keywords.



Low Intensity Search Example

All email communication from September 10, 2020 to October 1, 2020 between CDC Director Dr Redfield and Vice President Mike Pence.

- Mailbox – Dr Redfield’s mailbox(oxx0@cdc.gov)
- Date - 09/10/2020 – 10/01/2020

Moderate Intensity Search

- Custodian mailboxes defined.
- Participants may be known.
- Search on group mailboxes
- Search for phrase(s).
- Search for combination of keywords
- Extended date range.



Moderate Intensity Search Example

I'd like to make a public records request for any records pertaining to the CDC investigation about the circumstances in which a man died of COVID-19 in a commercial bus and why the commercial bus was transporting passengers with COVID-19. The bus, belonging the company El Tornado, departed on Oct. 13, 2020 form a Seneca Foods labor camp in Gillett, Wisconsin. It made a stop in Chicago, where it picked passengers not affiliated with the company, and then made another stop in Effingham, Illinois, that same day. At least another passenger of the bus, Jose Angel Rosas, would later die of COVID-19 in Laredo

- **Mailbox** – Not Defined. Program provided the custodian details.
- **Date** - October 18, 2020 To January 18, 2021
- **Keywords** – None



Moderate Intensity Search Outcome

Search Query - (**Bus** AND (**Milwaukee OR Effingham OR Dallas OR Laredo OR "San Antonio" OR McAllen OR "Hidalgo County" OR "Oconto County"**)) OR (**Motorcoach OR Seneca OR Gillett OR "Land conveyance"**) (c:c)(date=2020-10-13..2021-01-18)

Doc ID	Pages	Sent/Last Modified Date	Name	Dup	Cont	Size
520235	2	10/14/2020 5:53:15 PM	RE: Urgent situation - bus in US en route to Mexico with pro...	0		46 KB
520236	1	10/28/2020 11:00:33 AM	FW: CSTE Protocol Training - land conveyance reporting is b...	0		36 KB
520237	1	11/10/2020 10:49:17 AM	FW: Update on COVID-19 bus contact investigation and FDA...	0		37 KB
520238	1	11/1/2020 11:09:42 PM	RE: FOR REVIEW: Bus Investigation - documents for clearance	0		36 KB
520241	6	12/24/2020 3:08:26 PM	Fwd: Bus investigation IM slides for 12/29	0		801 KB
520242	21	12/9/2020 5:32:18 PM	Fwd: Bus investigation - case finding	0		3 MB
520244	1	11/9/2020 6:13:46 PM	FW: Update on COVID-19 bus contact investigation and FDA...	0		43 KB
520248	1	10/28/2020 11:49:31 AM	Re: QuarTET update	0		38 KB
520255	1	10/24/2020 6:31:23 PM	Re: CI for Bus land passenger from Mexico	0		51 KB
520256	5	1/4/2021 4:33:31 PM	FW: ABSTRACT REVIEW - Final Version and version with com...	0		156 KB
520257	1	10/15/2020 2:47:38 PM	Review notification write up	0		70 KB
520258	NA	10/15/2020 4:01:10 PM	Notification of Contacts exposed to COVID-19 case- 257519	0		1 MB
520259	4	11/10/2020 8:02:54 AM	RE: Update on COVID-19 bus contact investigation and FDA...	0		137 KB
520263	2	10/20/2020 12:41:29 PM	RE: Bus Travel Incident/Follow-up Discussion with Wisconsin	0		86 KB
520264	2	10/28/2020 11:50:32 AM	RE: QuarTET update	0		79 KB
520265	NA	10/28/2020 2:01:29 PM	FW: Addendum to Notification of Contacts exposed to COVI...	0		279 KB
520270	4	1/8/2021 8:53:19 AM	RE: Meeting to discuss motorcoach bus industry?	0		126 KB
520271	2	10/14/2020 4:57:11 PM	FW: QARS 257519 and QARS 257538	0		111 KB
520273	3	10/14/2020 9:20:29 PM	Possible DNB with AK QARS 257519 and QARS 257538	0		292 KB

High Intensity Search

- Custodian mailbox(s) defined.
- Too many custodian mailboxes.
- Participants not known.
- No Keywords to search.
- Combination of Keywords(OR & AND).
- Too many records.
- Too many attachments.



High Intensity Search Example

Copies of all responsive records held by the Centers for Disease Control and Prevention (CDC) — including procedures, internal memoranda, and emails — related to public guidance on wearing face masks to slow the spread of COVID-19.

- **Mailbox** – *Defined*
- **Date** - *January 1, 2020 and April 5, 2020*
- **Keywords** – *face mask(s), face covering(s), respirator(s), N95*

High Intensity Search Outcome

Search Query - ("mask*" OR ("face" AND "mask*") OR ("face" AND "cover*") OR "N95" OR "respirator*")

(from=ooo1@cdc.gov)(from=aaa1@cdc.gov)(from=mmm2@cdc.gov)(from=eee0@cdc.gov)(from=sss6@cdc.gov)

(date=2020-01-01..2021-04)

Location type	Keywords	Record Count
Exchange mailboxes	"mask*"	12,775
Exchange mailboxes	"respirator*"	12,155
Exchange mailboxes	"face" AND "cover*"	6,684
Exchange mailboxes	"face" AND "mask*"	5,722
Exchange mailboxes	"N95"	2,369
Total		22,877

ES Issues

- Broad Scope
- High Record Count
- Average Data Quality



ES Issues - Broad Scope

- Too many Keywords
- Generic keyword(s)
- Too many custodian mailboxes
- Extended date range



ES Issues - High Record Count

- High data volumes.
- Difficulty in identifying CORE data.



ES Issues - Average Data Quality

- Keywords missing in email content.
- Keywords found in attachments.
- Keywords found in email signature.



Improve ES

- Well defined Scope.
- Limit Keywords.
- Avoid Generic Keywords.
- Limit Custodians.
- Reduce Timespan.



Questions ?

Srinath Tutukuri – CDC FOIA Project Manager

OHF8@cdc.gov

404-718-6159

Roger Andoh – CDC/ATSDR FOIA Officer

MHU9@cdc.gov

770-488-6277

Bruno Viana – CDC/ATSDR Deputy FOIA Officer

CQY8@cdc.gov

770-488-6246

