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

**Analyze Search**
- Keywords
- Custodian Mailboxes
- Time Span

**Perform Search**
- MS 365 Compliance
- MS Outlook
- NUIX

**Eliminate Clutter**
- Subscriptions
- NewLetters
- Security Notifications

**Rerun Search**
- MS 365 Compliance
- MS Outlook
- NUIX

**Present Data**
- Enterprise Tools
- Shared drives

**Prepare Data**
- Dedupe
- Containment

**Export Data**
- .pdf, .doc
- .msg, .eml
- .pst

**Analyze Results**
- Results quality
- Capture metrics
- Inform Analyst
ES Tools

- Microsoft 365 Compliance
- Microsoft Outlook
- FOIA Express
- NUIX
ES Dedupe & Containment

- No of Input Files: 903
- No of Documents Failed Indexing: 3
- No of Documents After Indexing: 903
- All Duplicate Documents: 0
- All Contained Documents: 273
- No of Unique Documents: 630
- No of Representative Documents: 168
- No of Documents Moved to DM: 0
- No of Documents Not Moved to DM: 903
- No of Responsive Documents: 0
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 from 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 Sells OR Laredo OR "San Antonio" OR Milwaukee OR "Hidalgo County" OR "Oconto County"])) OR ([Motorcoach OR Seneca OR Gillett OR "Land conveyance"])(c:c)(date=2020-10-13..2021-01-18)

<table>
<thead>
<tr>
<th>Doc ID</th>
<th>Pages</th>
<th>Sent/Last Modified Date</th>
<th>Name</th>
<th>Dup</th>
<th>Cont</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>520235</td>
<td>2</td>
<td>10/14/2020 5:53:15 PM</td>
<td>RE: Urgent situation - bus in US en route to Mexico with prod...</td>
<td>0</td>
<td></td>
<td>46 KB</td>
</tr>
<tr>
<td>520236</td>
<td>1</td>
<td>10/28/2020 11:00:33 AM</td>
<td>FW: CSTF Protocol Training - land conveyance reporting is b...</td>
<td>0</td>
<td></td>
<td>36 KB</td>
</tr>
<tr>
<td>520237</td>
<td>1</td>
<td>11/10/2020 10:49:17 AM</td>
<td>FW: Update on COVID-19 bus contact investigation and FDA...</td>
<td>0</td>
<td></td>
<td>37 KB</td>
</tr>
<tr>
<td>520238</td>
<td>1</td>
<td>11/1/2020 11:09:42 PM</td>
<td>RE: REVIEW: Bus Investigation - documents for clearance</td>
<td>0</td>
<td></td>
<td>36 KB</td>
</tr>
<tr>
<td>520241</td>
<td>6</td>
<td>12/24/2020 3:08:26 PM</td>
<td>Fwd: Bus investigation IM slides for 12/29</td>
<td>0</td>
<td></td>
<td>801 KB</td>
</tr>
<tr>
<td>520242</td>
<td>21</td>
<td>12/9/2020 5:32:18 PM</td>
<td>Fwd: Bus investigation - case finding</td>
<td>0</td>
<td></td>
<td>3 MB</td>
</tr>
<tr>
<td>520244</td>
<td>1</td>
<td>11/9/2020 6:13:46 PM</td>
<td>FW: Update on COVID-19 bus contact investigation and FDA...</td>
<td>0</td>
<td></td>
<td>43 KB</td>
</tr>
<tr>
<td>520248</td>
<td>1</td>
<td>10/28/2020 11:49:31 AM</td>
<td>Re: QuatTET update</td>
<td>0</td>
<td></td>
<td>38 KB</td>
</tr>
<tr>
<td>520255</td>
<td>1</td>
<td>10/24/2020 6:31:23 PM</td>
<td>Re: Cl for Bus land passenger from Mexico</td>
<td>0</td>
<td></td>
<td>51 KB</td>
</tr>
<tr>
<td>520256</td>
<td>5</td>
<td>1/4/2021 4:33:31 PM</td>
<td>FW: ABSTRACT REVIEW - Final Version and version with com...</td>
<td>0</td>
<td></td>
<td>156 KB</td>
</tr>
<tr>
<td>520257</td>
<td>1</td>
<td>10/15/2020 2:47:38 PM</td>
<td>Review notification write up</td>
<td>0</td>
<td></td>
<td>70 KB</td>
</tr>
<tr>
<td>520258</td>
<td>NA</td>
<td>10/15/2020 4:01:10 PM</td>
<td>Notification of Contacts exposed to COVID-19 case- 257519</td>
<td>0</td>
<td></td>
<td>1 MB</td>
</tr>
<tr>
<td>520259</td>
<td>4</td>
<td>11/10/2020 8:02:54 AM</td>
<td>RE: Update on COVID-19 bus contact investigation and FDA...</td>
<td>0</td>
<td></td>
<td>137 KB</td>
</tr>
<tr>
<td>520262</td>
<td>2</td>
<td>10/20/2020 12:41:29 PM</td>
<td>RE: Bus Travel Incident/Follow-up Discussion with Wisconsin</td>
<td>0</td>
<td></td>
<td>86 KB</td>
</tr>
<tr>
<td>520264</td>
<td>2</td>
<td>10/28/2020 11:50:32 AM</td>
<td>RE: QuatTET update</td>
<td>0</td>
<td></td>
<td>79 KB</td>
</tr>
<tr>
<td>520265</td>
<td>NA</td>
<td>10/28/2020 2:01:29 PM</td>
<td>FW: Addendum to Notification of Contacts exposed to COVI...</td>
<td>0</td>
<td></td>
<td>279 KB</td>
</tr>
<tr>
<td>520270</td>
<td>4</td>
<td>1/8/2021 8:53:19 AM</td>
<td>RE: Meeting to discuss motorcoach bus industry?</td>
<td>0</td>
<td></td>
<td>126 KB</td>
</tr>
<tr>
<td>520271</td>
<td>2</td>
<td>10/14/2020 4:57:11 PM</td>
<td>FW: OARS 257519 and OARS 257538</td>
<td>0</td>
<td></td>
<td>111 KB</td>
</tr>
<tr>
<td>520272</td>
<td>3</td>
<td>10/14/2020 9:20:29 PM</td>
<td>Possible DNR with AK OARS 257519 and OARS 257538</td>
<td>0</td>
<td></td>
<td>292 KB</td>
</tr>
</tbody>
</table>

---

Centers for Disease Control and Prevention

---

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

<table>
<thead>
<tr>
<th>Location type</th>
<th>Keywords</th>
<th>Record Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange mailboxes</td>
<td>&quot;mask*&quot;</td>
<td>12,775</td>
</tr>
<tr>
<td>Exchange mailboxes</td>
<td>&quot;respirator*&quot;</td>
<td>12,155</td>
</tr>
<tr>
<td>Exchange mailboxes</td>
<td>&quot;face&quot; AND &quot;cover*&quot;</td>
<td>6,684</td>
</tr>
<tr>
<td>Exchange mailboxes</td>
<td>&quot;face&quot; AND &quot;mask*&quot;</td>
<td>5,722</td>
</tr>
<tr>
<td>Exchange mailboxes</td>
<td>&quot;N95&quot;</td>
<td>2,369</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td><strong>22,877</strong></td>
</tr>
</tbody>
</table>
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