

Summary for stakeholders

Mohammed Beheiry · FEC Contributions Analysis · 2017–2020

Key findings

Under the WinRed-exclusion scenario, itemized contribution records total about USD 15.25 billion. Individuals account for about 83% of that amount. Contributions peaked in October 2020, just before the presidential election. Committee contributions are concentrated: the top 10% of candidates received about 60% of the attributed amount, while the median candidate received about USD 16,000.

- Contributions roughly doubled from the 2018 cycle (about USD 5.09 billion) to the 2020 cycle (about USD 10.16 billion). Individual contributions more than doubled, while committee contributions grew by about 40%.
- The reported-record measure is about USD 1.55 billion higher; the difference is the WinRed-reported records excluded from this scenario. One candidate's USD 49.6 million committee total falls to about USD 1.2 million without them.
- The source needed cleanup before counting. The 2020 individual table repeats records up to five times (195.5 million rows for 88.8 million unique IDs), and the table labeled 2020 operating expenditures contains 2017–2018 reporting years, so I did not use it.
- Small donations are undercounted. The data contains itemized records, and donors who give USD 200 or less to a committee generally do not have to be itemized.

These figures exclude WinRed conduit reporting to reduce overlap. They are not a verified count of unique donations. ActBlue and WinRed report differently, so comparisons between parties require care. The technical report shows the alternative reported-record measure.

Key numbers

Result	WinRed-exclusion scenario
Individual contributions	57.5 million records · USD 12.67 billion
Committee contributions	653 thousand records · USD 2.58 billion
Busiest month	October 2020 · up 4.4% from September
Candidates receiving committee contributions	2,603
Median / average received per candidate	USD 16,000 / USD 405,000
Top 10 businesses per state	Preliminary version: 4 hours; validated version: estimated 3–4 working days

Decisions to confirm

Agree on the contribution measure to use for future reporting. For the business ranking, confirm whether to use the payee's or committee's state, gross payments or payments net of refunds, whether to combine subsidiaries with parent companies, and how to treat payment processors and other intermediaries.

Next steps and data needed

Next step	Planning estimate	Data and availability
Confirm ranking scope	30 minutes	Stakeholder decisions needed
Build a preliminary ranking	4 hours, including scope	Expenditure tables available; coverage still needs validation
Validate businesses and intermediary payments	3–4 working days total	Available records plus manual review; further evidence may be needed
Check completeness of 2018 expenses	Half a day initially	Official bulk files: not included in this analysis; need to be obtained
Compare party and incumbent groups	1 hour initially	Candidate tables available; attribution coverage needs checking
Assess unitemized contributions	Estimate after source review	Committee summary reports: not included in this analysis; need to be obtained

Estimates depend on source coverage and the number of unresolved businesses. They are planning allowances, not measured delivery times, and should not be added together. Summary reports could extend aggregate coverage but would not provide a complete transaction-level list of small donations.

Technical report

The stakeholder summary uses the WinRed-exclusion scenario. Q1 and Q2 below show the original reported-record results and explain how that scenario changes them. Both use the 2018 and 2020 cycles, covering January 2017 through December 2020.

Answers by question

Question	Reported-record measure	WinRed-exclusion scenario
1a. Individual contributions	79,847,961 records	57,464,005 records
1b. Committee contributions	1,750,504 records	652,869 records
1c. Individual / committee amounts	USD 14.08B / USD 2.72B	USD 12.67B / USD 2.58B
1d. Peak month / growth from prior month	October 2020 / 12.50%	October 2020 / 4.44%
2a. Candidate IDs receiving committee contributions	2,614	2,603
2b. Median / average per candidate ID	USD 15,800 / USD 428,635	USD 15,750 / USD 404,860
2b. Share received by the top 10% of candidate IDs	61.91%	59.61%

Amounts combine both cycles. Medians are approximate; averages and amounts are rounded. The scenario reduces overlapping reporting but is not a verified count of unique donations.

[Q1: sources](#) · [Q2: candidates](#) · [Q3: analysis plan](#) · [QA and model](#) · [Conduit comparison](#) · [Scope and next steps](#) · [Appendix](#)

Q1. Contribution counts and amounts

Source	Cycle	Contribution records	Gross reported amount
Individuals	2018	15,088,198	USD 4,021,953,507
Individuals	2020	64,759,763	USD 10,058,325,757
Individuals	Both cycles	79,847,961	USD 14,080,279,264
Committees	2018	334,553	USD 1,067,216,231
Committees	2020	1,415,951	USD 1,653,263,736
Committees	Both cycles	1,750,504	USD 2,720,479,967

SQL: 16_model_committee_q1 — answer statement

SQL: 17_model_individual_q1 — answer statement

Each included record ID is unique within its source and cycle. The counts describe transactions, not distinct donors. Committee activity includes payments to candidate committees and other registered committees.

The table uses the reported-record measure. Under the WinRed-exclusion scenario, individual contributions fall to 57,464,005 records totaling USD 12,667,245,189 and committee contributions fall to 652,869 records totaling USD 2,579,862,501. The stakeholder summary uses these figures. I explain the scenario and its limits in [Intermediary reporting and overlap](#).

Peak month

October 2020 recorded USD 2,017,565,781 in combined activity, compared with USD 1,793,420,123 in September. The month-over-month increase is 12.50%.

SQL: 18_model_monthly_contributions — peak-month statement

I aggregated the records by calendar month and used a complete monthly calendar before calculating the prior-month comparison. The monthly table reconciles exactly to the source counts and amounts, with 48 distinct months per source. It also keeps chart queries small.

The peak and growth rate apply to the reported-record definition. Excluding WinRed's 15E conduit receipts and its outgoing committee contributions leaves October 2020 as the peak, at USD 1,636,196,210 versus USD 1,566,597,035 in September, an increase of 4.44%. This tests the effect of the reporting choice; it is not a fully deduplicated donation estimate.

Source queries: 16_model_committee_q1, 17_model_individual_q1, 18_model_monthly_contributions, 19_qa_monthly_reconciliation and 21_monthly_chart_data.

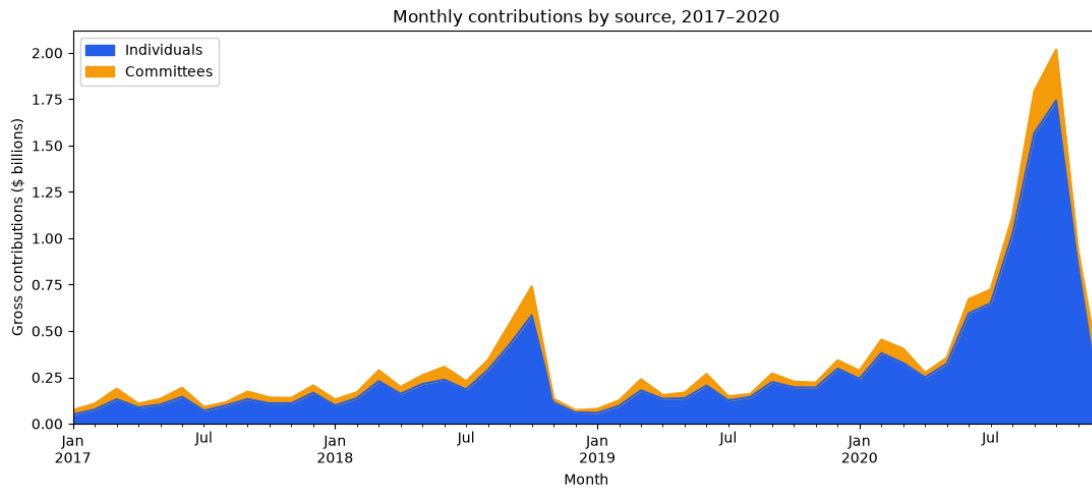
Definitions and caveats

I used gross positive amounts so the contribution measure has a consistent meaning across sources. Refunds, zero amounts, nonblank memo codes, and records with missing or out-of-cycle dates are excluded. Both memo codes X and Y are excluded by the same rule. Amounts are cast to NUMERIC for aggregation.

Individual receipts require entity type IND and transaction types 10, 15, 15E, 30, 30E, 31, 31E, 32, or 32E. This excludes candidate self-contributions, loans and intermediary disbursements. The source contains itemized records; it is not a complete measure of all donations.

Committee contributions use outgoing transaction types 24K and 24Z. For Q1, I used other_committee_tx_2018 and other_committee_tx_2020. For Q2, I used committee_contributions_2018 and committee_contributions_2020 because they contain candidate attribution. I did not add the two committee sources together.

The [FEC transaction code definitions](#) informed these filters. The unit of analysis is a qualifying record within a cycle; candidate statistics combine records by candidate ID across both cycles.



Individual receipts make up most of the reported amount at the October 2020 peak. The stacked areas show the contribution of each source to the combined monthly total. Source: 21_monthly_chart_data.csv.

SQL: 21_monthly_chart_data

Q2. Candidate coverage and distribution

Qualifying committee records identify 2,614 candidate IDs and USD 1,120,452,295 in attributed contributions across both cycles. The median total per candidate ID is approximately USD 15,800, compared with a mean of USD 428,635. The top 262 IDs receive 61.91% of the attributed amount.

Measure	Total per candidate ID	What it means
Minimum	USD 8	The lowest total received by a candidate ID in this analysis.
25% approximately	USD 3,500	About 25% of candidate IDs received this amount or less.
Median (50% approximately)	USD 15,800	About half received this amount or less, and half received this amount or more.
Average (mean)	USD 428,635	Total attributed contributions divided by the number of candidate IDs. Large totals pull the average up.
75% approximately	USD 369,013	About 75% of candidate IDs received this amount or less.
90% approximately	USD 1,494,038	About 90% of candidate IDs received this amount or less.
99% approximately	USD 3,946,544	About 99% of candidate IDs received this amount or less.
Maximum	USD 49,608,977	The highest total received by a candidate ID in this analysis.
Standard deviation	USD 1,271,873	Measures how spread out the totals are around the average, using all candidate IDs in this analysis. A larger value means more variation.

SQL: 12_q2_candidate_summary

Definitions and caveats

I aggregated by candidate ID before calculating the distribution. The large difference between the mean and median reflects concentration at the upper end.

Populated IDs remain in the analysis even where the same-cycle candidate reference has no match: 138 IDs in 2018 and 125 in 2020. The result is therefore a count of reported candidate IDs, not a verified count of unique people. Another 2,938 qualifying records, totaling USD 3,098,774, have no candidate ID and are excluded from the distribution.

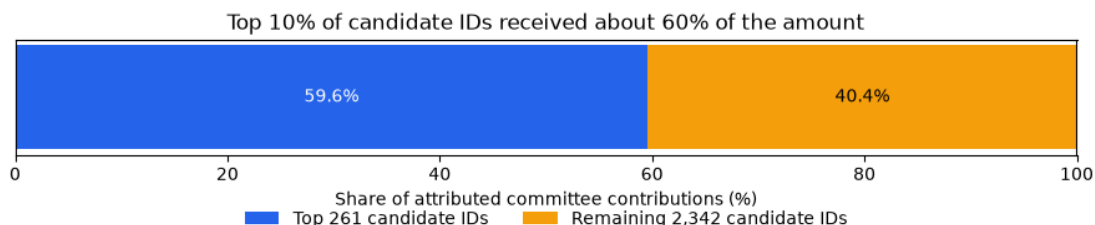
Largest candidate total

The USD 49,608,977 maximum belongs to Donald Trump's candidate ID, P80001571. Of that amount, USD 48,407,633 was reported by WinRed, C00694323, in 2020: 257,625 records, all coded 24K, with a largest record of USD 5,600. WinRed accounts for 97.58% of this candidate's total. The full breakdown reconciles to the original maximum.

The WinRed records identify recipient committee C00580100, use entity type CCM, and have no memo text. They establish the reported amount and recipient attribution, but do not identify the original funding source of each payment. I have retained them in the stated committee-record measure. I would not describe the result as USD 49.6 million of independently funded PAC support.

Excluding WinRed from the committee-source measure reduces coverage from 2,614 to 2,603 candidate IDs and the attributed amount to USD 1,053,851,132. The approximate median is USD 15,750 and the mean is USD 404,860. The largest total becomes USD 6,283,681, for candidate ID S4GA11285; Trump's remaining total is USD 1,201,344. The top 261 IDs still account for 59.61% of the amount. Concentration therefore persists, but the original maximum is heavily affected by conduit reporting. The stakeholder summary uses these figures. Other intermediaries may still be included.

Candidate and committee identities were checked against the [FEC candidate record](#) and [FEC committee record](#). Source queries: 11_model_committee_staging, 12_q2_candidate_summary, 22_qa_largest_candidate, 24_qa_winred_candidate_reporting and 27_qa_winred_sensitivity.



WinRed-exclusion scenario, both cycles. The top group contains 261 of 2,603 candidate IDs, rounding 10% up to a whole candidate. Source: query 27. This shows concentration; the table above describes the original distribution.

Q3. Analysis plan: top 10 business payees by state

A preliminary ranking is feasible within four hours. A validated ranking is not: I estimate about 26 hours, or three to four working days. The aggregation itself is straightforward. Business identification, name standardization and payments through intermediaries take most of the time, and all three depend on manual review.

Source	Records	Dates in cycle	Missing dates	Reporting years
opex_2018	1,800,080	0	1,800,080	2017–2018
opex_2020	1,800,080	0	1,800,080	2017–2018
oppexp18	1,800,000	1,795,535	252	2017–2018
oppexp20	2,315,382	2,310,905	15	2019–2020

The query below produced the source assessment. Q3 asks for an analysis plan; I have not produced a business-payee ranking.

SQL: 20_qa_operating_expenditures

Approach and limitations

I would start with oppexp18 and oppexp20. Their reporting years match the requested cycles, and most dates fall within those periods. I ruled out opex_2020 for the 2020 analysis: its reporting years are 2017–2018 and every transaction date is missing. Its summary statistics also match opex_2018.

The alternative sources have missing names and states, negative amounts and implausible dates. oppexp18 contains exactly 1,800,000 records. That is a reason to verify coverage, although that count alone is not proof of truncation. opex_2018 has 80 more records and USD 377,626 more than oppexp18, so the two 2018 versions differ even apart from the missing dates.

I would first check record IDs, dates, memo entries and adjustments. Records without usable names or states would be accounted for separately. I would then standardize payee names and review likely matches among the largest recipients, using address details where available. Similar names alone would not justify merging businesses.

Intermediary payments need particular care. A processor or payroll provider may be the reported payee without being the final recipient. Supporting itemization must not be counted again as an additional payment. After applying those rules, I would aggregate accepted amounts by business and state, rank within each state, and review the leading results.

Questions for the stakeholder

I would confirm these before building the ranking:

- Should "state" mean the payee's reported address or the paying committee's state?
- Should the ranking use gross payments, or payments net of refunds and adjustments?
- Should a business be combined with its subsidiaries or parent company?
- Does a payment processor, payroll provider or conduit count as the business receiving the money?
- Should the two cycles be combined or ranked separately?

Until those are settled, my starting assumption would be total gross payments across both cycles, grouped by the payee's reported state.

Time estimate

Step	Preliminary (hours)	Validated (hours)	What takes the extra time
Confirm scope with the stakeholder	0.25	0.5	Agreeing on the questions above and a few edge cases
Record checks: IDs, dates, memo entries, negative amounts and coverage	0.75	3	Comparing oppexp18 with opex_2018 by record ID to confirm whether oppexp18 is complete, and deciding how to treat 38,481 negative amounts and dates such as 9201-10-02
Identify business payees	0.5	4	Entity type ORG covers any organization, including government agencies and nonprofits, and 167,828 oppexp18 records have no entity type
Standardize payee names	1	8	Basic cleanup is quick, but merging spelling variants of the same business needs address checks and a review of each proposed match
Intermediary payments	0.5	6	Credit card, payroll and processor payments have to be traced to their supporting itemization so the money is not credited to the processor or counted twice
Aggregate and rank within each state	0.25	0.5	The ranking query itself is simple
Review the results and document exclusions	0.75	4	Reviewing the top 10 in every state, not just the largest ones, and walking the stakeholder through unresolved cases
Total		4	about 26

The four-hour version would contain the preliminary ranking, the SQL, the definitions and the value of excluded or unresolved records. I would spend the review time on cases large enough to change a top 10, and I would label the result as a ranking of identified business payees in the available records, not a complete ranking of all businesses.

Source query: 20_qa_operating_expenditures.

Source validation and modeling decisions

I compared table coverage before choosing sources. Similar names and matching totals were not sufficient evidence that two versions contained the same records.

Check	Finding	Decision
Individual record IDs	2018 has 21,628,973 unique IDs. 2020 has 195,482,945 rows and 88,823,302 unique IDs. Neither has missing IDs.	Removed repeated copies using stable analysis fields within each cycle.
Repeated-ID differences	104,323 IDs differ in amendment indicator; the other checked fields agree.	Kept the agreed analysis fields without choosing an amendment indicator.
Amount and date profile	Positive refunds, memo records, non-individual entities and invalid dates occur in the sources.	Applied type, entity, memo, sign and date filters together.
Alternative 2020 candidate sources	885,084 shared IDs; 2,745 descriptive-only and 4,081 pas220-only IDs.	Kept the versions separate.
Candidate attribution	164 contribution records, worth USD 443,400 before scope filters, have candidate IDs in the descriptive source but not pas220. One USD 1,000 contribution has different populated IDs.	Selected the descriptive source for Q2; attribution differences remain disclosed.
Committee-source overlap	275,881 qualifying IDs overlap in 2018 and 734,239 in 2020; amount, date and committee agree.	Used the broader source alone for Q1.
Unmatched committee records	25 qualifying candidate records worth USD 39,000 are absent from the broader 2020 source.	Disclosed the coverage difference and did not append the unmatched records.
Conduit reporting	ActBlue's side of earmarked donations is coded 24T and falls outside the filters; WinRed's side is coded 15E and passes them. 1,916,040 one-to-one field matches between WinRed and recipient reports; other matches are ambiguous.	Used a WinRed-exclusion scenario in the stakeholder summary and kept the reported-record measure alongside it. The scenario reduces overlap but does not prove each donation is counted once.
Monthly reconciliation	Counts and amounts match for both sources; each has 48 distinct monthly rows.	Used the reconciled monthly aggregate for charts and comparisons.

Deduplicating record IDs does not identify every repeated underlying donation. Likewise, retaining a populated candidate ID does not validate every attribution. These are separate questions from whether the SQL preserves the selected records.

Staging rows retain eligibility flags. Query 29 saves the deduplicated individual rows in a table clustered by cycle and committee; the staging view now reads that snapshot. The monthly aggregate remains the reporting layer for charts.

The repeated-ID comparison processed 14.73 GB in 13.821 seconds. I used job information to monitor execution cost and runtime; the reconciliation checks provide the evidence for aggregation accuracy.

Source queries: 01_source_inventory through 10_qa_committee_profile; 13_qa_other_committee_profile through 19_qa_monthly_reconciliation.

Model	Grain	Purpose	Build query
stg_committee_contributions	Source record and cycle	Candidate attribution and Q2 eligibility	11
stg_committee_q1	Source record and cycle	Broader committee activity and Q1 eligibility	16
individual_records_cached / stg_individual_q1	Selected record fields and cycle	Saved deduplicated rows, including excluded records and eligibility flags	17, 29
fact_contributions	Source type, cycle and record ID	Q1 amounts, reporting and recipient committees, candidate attribution where available, and reporting-platform flag	18, 29
dim_committee	Committee ID and cycle	Committee names, types, party and state	29
dim_candidate	Candidate ID and cycle	Candidate names, party, office and incumbent status	29
agg_contributions_monthly	Month and source type	Materialized counts and gross amounts	18

The views preserve the selected source fields, not every original column. The monthly table supports repeated reporting without rescanning the individual source. Adding cycle to the record key keeps election-cycle extracts separate.

The dimensions have one row per ID and cycle. Multiple source labels are retained together rather than silently choosing one. Candidate attribution is attached only to matching committee records; individual receipts are not assigned to candidates through an assumed committee relationship. The platform flag identifies the reporting committee, not the original donor. Query 29 checks the richer fact view against the existing monthly counts and amounts before it is used for reporting.

The snapshot is a fixed extract, not an automatic refresh pipeline. Query 29 rebuilds it directly from the original sources. Query 18 can rebuild the monthly table; query 29 should run last to restore the enriched fact view. This keeps the original analysis reproducible while supporting repeated investigation without repeating the large deduplication.

For conduit-reported individual receipts, `recipient_committee_id` is the reporting conduit's ID, not necessarily the final recipient committee.

Intermediary reporting and overlap

The committee-source comparison found 1,010,120 shared qualifying record IDs. Using one source for Q1 avoids adding those records twice. I also checked a different issue: the same donation appearing under separate IDs in a conduit's report and a recipient's report.

ActBlue and WinRed report their side of an earmarked donation with different transaction codes, so my filters treat them differently.

2020 individual records	Reported by the conduit	Reported by the recipient	Counted under my filters
ActBlue	20,733,195 records coded 24T, USD 2,140,427,718	18,626,946 records coded 15E, USD 1,801,148,249	Recipient side only
WinRed	22,383,956 records coded 15E, USD 1,413,034,075	4,125,031 records coded 15E, USD 584,456,022	Both sides

My transaction filter excludes 24T, so ActBlue donations are counted through the recipient's report. WinRed codes its own receipts 15E, the same as the recipient, so both sides pass the transaction and memo filters. ActBlue follows the same pattern in 2018, and WinRed has no records in the 2018 data. A blank memo code alone does not resolve conduit reporting.

WinRed's other_id field identifies an intended recipient on all but 83 of its included records. I compared that ID with the receiving committee's ID, together with the donor's name, five-digit ZIP code, date and amount. Names were trimmed and converted to uppercase. The comparison returned 1,916,040 groups with exactly one record on each side, totaling USD 341,092,327 per side. Another 177,252 groups contained multiple records on one or both sides. Many records had no match under these fields.

These matches are strong evidence of repeated reporting, but the matching fields are not a unique donation identifier. I did not automatically delete matching records or treat unmatched records as proof of separate donations. The [FEC's reporting guidance](#) explains that recipients disclose the original contributor and the date the conduit received the contribution. [WinRed's conduit guidance](#) also describes reporting by both parties.

To measure the effect on the answers, I excluded WinRed-reported 15E receipts and WinRed's outgoing committee contributions while retaining recipient-reported individual receipts. All other filters stayed the same.

Measure, both cycles	Reported-record measure	WinRed-exclusion scenario
Individual records	79,847,961	57,464,005
Individual amount	USD 14,080,279,264	USD 12,667,245,189
Committee records	1,750,504	652,869
Committee amount	USD 2,720,479,967	USD 2,579,862,501
Peak month	October 2020	October 2020
Peak-month growth	12.50%	4.44%
Candidate IDs in Q2	2,614	2,603
Largest candidate total	USD 49,608,977	USD 6,283,681

The timing of the peak and the concentration of candidate receipts hold under this test. Their size is sensitive to how conduit activity is counted.

The summary uses the WinRed-exclusion scenario. It reduces overlapping reporting but also removes unmatched conduit records. A missing match can reflect differences in fields, reporting coverage or itemization; it is not proof that a recipient never reported the donation. The scenario is therefore not a fully deduplicated measure of original donor funding.

Committee-side follow-up

Query 28 checked reporting committees whose names contain ActBlue or WinRed in both committee sources. ActBlue has six included Q1 records totaling USD 237,000: five records in 2018 worth USD 37,000 and one in 2020 worth USD 200,000. It has no included Q2 records. WinRed has 1,097,635 included Q1 records worth USD 140,617,466 and 472,869 included Q2 records worth USD 66,674,728. The Q1 excluded amounts match query 27 exactly. The Q2 amount is USD 73,565 higher than the USD 66,601,163 removed in query 27, which restricts candidate totals to records with a populated candidate ID.

This closes the platform-coding check for the two sources. It does not trace the original funding behind ActBlue's six records or payments routed through other committees. I retained those records under the stated filters and made no further adjustment to the totals.

Source queries: 23_qa_intermediary_reporting through 28_qa_committee_platform_reporting. Results reviewed September 14, 2026.

SQL: 27_qa_winred_sensitivity

Scope and next steps

Scope

I focused the analysis on the three questions, source quality and checks that could change the answers. The report separates the findings from the detailed SQL and QA appendix so a reviewer can read the conclusions first and inspect the supporting work where needed.

Completed refinements

I materialized the deduplicated individual records so repeated analysis no longer reruns that step. I also added recipient and reporting-committee fields, a reporting-platform flag, candidate attribution where available, and committee and candidate dimensions. The updated model reconciles to the original totals. Q2 includes a concentration chart showing the share received by the top 10% of candidate IDs.

With more time

I would add a full Q2 distribution chart using candidate-level totals, with a log scale to make the smaller amounts visible alongside the largest recipients. The current chart shows concentration, not the shape of the full distribution. I would also automate refreshes and reconciliation for the saved individual table, and extend the conduit fields beyond the named reporting platforms to trace more of the path from the original donor to the recipient. Those additions would support recurring reporting and further analysis beyond this submission.