IT Collect "How To" Series

How to use the Data Quality Report

September 2021

Beta v0.1

General Services Administration

Office of Government-wide Policy

Document Change Record

Version Number	Date	Description
1.0	September 2021	Created IT How-To Data Quality Report document outlining IT Collect data reports.

What is the IT Collect Statistics service?

The IT Collect Statistics service is a GET-only, multi-report mechanism that displays data quality, data completeness, data staleness, and other metrics within reports to agencies. This series will focus in on the data quality report that allows users to better understand the composition of their investment data. Some portions of the data quality report contains their own sub-reports (Result Sets) that provide additional data visibility.

> *For IT Collect go-live in September, these are the 3 year values that will be used:*
> Budget Year (BY) = 2022
>
> Current Year (CY) = BY – 1 = 2021
>
> Previous Year (PY) = CY – 1 = 2020
>
> The active Time Period automatically assigned by IT Collect will match the Budget Year. The numbers 01, 02, 03, and 04 you see appended to the Time Period ID's in the below screenshots reference the different submission periods in order.

What is displayed in the Data Quality Report?

Stock Data Quality Report Endpoint

This collection reports on the quality/completeness of the data submitted to the IT Collect database for a given agency. These reports are designed to be viewed by agency end users via a data feed on the modernized IT Dashboard front-end. All reports apply to unretired UII's (isRetired = false) and non-Budget Authority Ledger Element Styles (ledgerElementStyle != "Budget Authority"). Below are the available filters, individual result sets, their subsets, and descriptions agencies will see in this report:

Filter	Description	Inputs
report	Required filter which allows users to designate which report to run	All available Result Sets outlined below
page	Specifies which page number a user wants to display. The first page will display 200 records, the second the next 200, and so on.	Integer value

Result Set	Subsets	Description	Fields Returned
missingCioRatings	N/A	*UII's that do not have CIO Ratings and have investmentType = 01 or 05*	agencyId (service) agency currentUII
missingContracts	N/A	*UII's that do not have Contracts and have investmentType = 01, 02, or 05*	agencyId (service) currentUII

missingProjects	N/A	*UII's that do not have Projects and have investmentType = 01*	agencyId (service) currentUII
activityStartDateTooEarly	N/A	*Activities with any start date before their parent project start date*	agencyId (activity) plannedStartDate projectedStartDate actualStartDate agencyId (service) currentUII AgencyId (project) StartDate (project)
activityEndDateTooLate	N/A	*Activities with any completion date following their parent project completion date*	agencyId (activity) plannedEndDate projectedEndDate actualEndDate agencyId (service) currentUII activityId (project) projectEndDate
newServicesWithCYPYFunding	N/A	*UII's with CY or PY funding AND changeInStatus = 10 (New Investment)*	year (ledger) amount (ledger) agencyId (ledger) currentUII agency agencyId (service)
eliminatedInvestmentsWithCYBYFunding	N/A	*UII's with a CY or BY funding AND changeInStatus = 06, 07, 08, or 09 (Eliminated Investments)*	year (ledger) amount (ledger) agencyId (ledger) currentUII agency AgencyId (service)
projectMissingActivities	N/A	*Projects with no activities*	agencyId (project) projectName agencyId (service) currentUII (service)
missingCostPools	missingAllCostPools	*UII's with no cost pools at all*	changeInStatus agencyId (service) currentUII
	missingPreviousYearCostPools	*Existing UIIs (changeInStatus != 10) missing PY cost pools*	agencyId (service) changeInStatus currentUII
	missingCurrentYearCostPools	*UII's that are not new or eliminated (changeInStatus != 10, 06, 07, 08, or 09) and are missing CY cost pools*	agencyId (service) changeInStatus currentUII

	missingBudgetYearCostPools	*UII's that are not eliminated (changeInStatus != 06, 07, 08, or 09) and are missing BY cost pools*	agencyId (service) changeInStatus currentUII
missingITTowers	missingAllITTowers	*UII's with no IT towers at all*	changeInStatus agencyId (service) currentUII
	missingPreviousYearITTowers	*Existng UIIs (changeInStatus != 10) missing PY IT towers*	agencyId (service) currentUII changeInStatus
	missingCurrentYearITTowers	*UII's that are not new or eliminated (changeInStatus != 10, 06, 07, 08, or 09) and are missing CY IT towers*	agencyId (service) currentUII changeInStatus
	missingBudgetYearITTowers	*UII's that are not eliminated (changeInStatus != 06, 07, 08, or 09) and are missing BY IT towers*	agencyId (service) currentUII changeInStatus
missingFundingSources	missingAllFundingSources	*UII's with no funding (henceforth derived from ledger records containing funding source data) at all*	changeInStatus agencyId currentUII
	missingPreviousYearFundingSources	*Existing UIIs (changeInStatus != 10) missing PY funding*	agencyId (service) currentUII changeInStatus
	missingCurrentYearFundingSources	*UII's that are not new or eliminated (changeInStatus != 10, 06, 07, 08, or 09) and are missing CY Funding Sources*	agencyId (service) currentUII changeInStatus
	missingBudgetYearFundingSources	*UII's that are not eliminated (changeInStatus != 06, 07, 08, or 09) and are missing BY Funding Sources*	agencyId (service) currentUII changeInStatus
cioRatingsInRedOverThreeMonths	N/A	*UII's with a red CIO rating (cioRating = 1*	title (service) currentUII agencyId (service)

		or 2) for more than 90 days	
cioRatingsInRedOverOneYear	N/A	*UII's with a red CIO rating (cioRatting = 1 or 2) for more than 365 days*	title (service) currentUII agencyId (service) date (CIO rating) rating (CIO rating) agencyId (CIO rating)
missingOperationalAnalyses	N/A	*UII's that are missing operational analysis data*	agencyId (service) currentUII
projectLifecycleLedgerExcess	N/A	*UII's where project lifecycle costs exceeds ledger lifecycle costs*	agencyId (service) currentUII title (service) projectLifecycleCost ledgerLifecycleCost
UIIsToBeRetired	N/A	*UII's that are eliminated (changeInStatus = 06, 07, 08, 09) and are not yet Retired (isRetired = false)*	changeInStatus currentUII title (service) agencyId (service)
activitiesOverOneYear	N/A	*Activities with a duration (planned/projected/ actual end date – planned/projected/a ctual start date) exceeding one year*	currentUII agencyId (service) title (service) agencyId (project) projectName agencyId (activity) activityName actualStartDate, plannedStartDate, ProjectStartDate actualEndDate, plannedEndDate, projectedEndDate
projectsMissingLowestLevelActivities	N/A	*UII's that do not contain a single activity where lowestLevel = true*	agencyId (service) title (service) currentUII agencyId (project) projectName
lowestLevelValidityCheck	N/A	*(1) UII's where an activity has lowestLevel = true but has a sibling activity that has a lower structureId OR (2) activity has*	agencyId (service) title currentUII agencyId (project) projectName agencyId activityName

		lowestLevel = false but does not have a sibling activity that has a lower structureId -- *Report can only be run on a single investment*	structureId providedLowestLevel expectedLowestLevel

❖ Here are two sample Result Sets users will see within the Data Quality Report. On the left is missingCioRatings and on the right is missingOperationalAnalyses. Users can expect all Result Sets under the Data Quality Report to be displayed similarly to these two examples:

```
{
  "@context": "/contexts/DataQualityReport",
  "@id": "/v1/data-quality-report",
  "@type": "hydra:Collection",
  "hydra:member": [
    {
      "_id": "01F3DCAH5RB6K20DXZZFTHRDSV",
      "agency": "018",
      "agencyId": "string",
      "currentUII": "018-000000283"
    },
    {
      "_id": "01F3DCGJR3XJ70XH2JFR4G6N63",
      "agency": "018",
      "agencyId": "sring",
      "currentUII": "018-000000293"
    },
    {
      "_id": "01F3DNVCECAY700V68SJ9WSHT3",
      "agency": "018",
      "agencyId": "1string",
      "currentUII": "018-000000124"
    }
  ],
  "hydra:totalItems": 3,
  "hydra:view": {
    "@id": "/v1/data-quality-report?report=missingCioRatings",
    "@type": "hydra:PartialCollectionView"
  },
  "hydra:search": {
    "@type": "hydra:IriTemplate",
    "hydra:template": "/v1/data-quality-report{?report}",
    "hydra:variableRepresentation": "BasicRepresentation",
    "hydra:mapping": [
      {
        "@type": "IriTemplateMapping",
        "variable": "report",
        "property": null,
        "required": true
      }
    ]
  }
}
```

```
{
  "@context": "/contexts/DataQualityReport",
  "@id": "/v1/data-quality-report",
  "@type": "hydra:Collection",
  "hydra:member": [
    {
      "agencyId": "01F3BES0NQHJRHN5XNY0Q3HZR7",
      "currentUII": "018-000000003"
    },
    {
      "agencyId": "01F3BES0NW1NTS3WCRNQW5YHXG",
      "currentUII": "018-000000004"
    },
    {
      "agencyId": "01F3BES0P2FQW0M31W91YB1J88",
      "currentUII": "018-000000005"
    }
  ],
  "hydra:totalItems": 3,
  "hydra:view": {
    "@id": "/v1/data-quality-report?report=missingOperationalAnalyses",
    "@type": "hydra:PartialCollectionView"
  },
  "hydra:search": {
    "@type": "hydra:IriTemplate",
    "hydra:template": "/v1/data-quality-report{?report}",
    "hydra:variableRepresentation": "BasicRepresentation",
    "hydra:mapping": [
      {
        "@type": "IriTemplateMapping",
        "variable": "report",
        "property": null,
        "required": true
      }
    ]
  }
}
```

TBM Discrepancy Report

This report displays a list of investments with either a cost pool or IT tower total for a given year that is greater than a plus or minus 5% delta than the funding amount total. All UII's displayed in

this report are not retired (isRetired != false) and the funding amount totals do not include budget authority ledgers (ledgerElementStyle != "Budget Authority"). Below are the available filters, individual result sets, and descriptions agencies will see in the report:

Filter	Description	Inputs
page	Specifies which page number a user wants to display. The first page will display 200 records, the second the next 200, and so on.	Integer value

Result Set	Description
funding_total	*Total Funding amount ($M) for a given year*
cost_pool_total	*Total Cost Pool amount ($M) for a given year*
cost_pool_deviation	*% Difference between Cost Pool total and Funding total for a given year*
it_tower_total	*Total IT Tower amount ($M) for a given year*
it_tower_deviation	*% Difference between IT Tower total and Funding total for a given year*

❖ Below is a sample report users will see for the TBM Discrepancy Report. Users can expect their report to be displayed in a similar way:

```
{
    "@context": "/contexts/TBMDiscrepancyReport",
    "@id": "/v1/tbm-discrepancy-report",
    "@type": "hydra:Collection",
    "hydra:member": [
        {
            "018-000000000": {
                "2019": {
                    "funding_total": 12.907,
                    "cost_pool_total": 27.967,
                    "cost_pool_deviation": 1.166809,
                    "it_tower_total": 0,
                    "it_tower_deviation": -1
                },
                "2020": {
                    "funding_total": 5.57,
                    "cost_pool_total": 0.6970000000000001,
                    "cost_pool_deviation": -0.874865,
                    "it_tower_total": 0,
                    "it_tower_deviation": -1
                },
                "2021": {
                    "funding_total": 50.14,
                    "cost_pool_total": 17.869,
                    "cost_pool_deviation": -0.643618,
                    "it_tower_total": 58.183,
                    "it_tower_deviation": 0.160411
                }
            }
        }
    ],
    "hydra:totalItems": 1
}
```

Portfolio Report

This report displays all non-retired (isRetired = false) investment-level UII's for each time period. This report is intended for agencies to see how many Services they have submitted across their portfolio. Here is an example of this report and available filters:

Filter	Description	Inputs
report	Only one result set is available for the Portfolio Report	PortfolioReport
page	Specifies which page number a user wants to display. The first page will display 200 records, the second the next 200, and so on.	Integer value

```json
{
    "portfolioReport": {
        "TimePeriod20211": {
            "018": [
                "018-000000000",
                "018-000000001",
                "018-000000002",
                "018-000000003",
                "018-000000004",
                "018-000000005"
            ]
        },
        "TimePeriod20212": {
            "018": [
                "018-000000000",
                "018-000000001",
                "018-000000002",
                "018-000000003",
                "018-000000004",
                "018-000000005"
            ]
        }
    }
}
```