

# Introductions

Monday, May 1, 2023 2:07 PM

You have been booked as follows:

Customer: UHC Medicare & Retirement  
Project: MRIS DQ Enablement  
Project ID: 119394  
Requester: Evelyn A Smith

for the period

Starting on: 05/08/23  
Ending on: 07/17/23

Booking type: 1-Sold  
Booked: 100%

DQaaS: <https://hcp.uhg.com/products/dqaaS>

MRIS DQ - <https://teams.microsoft.com/l/team/19%3a2ohHhyZ2zctV6-lVwRNXevNmDcIGJRMEd0sZhRM7Qrl1%40thread.tacv2/conversations?groupId=6e770f14-ed63-47cf-b260-b4a1f75cd9d2&tenantId=db05faca-c82a-4b9d-b9c5-0f64b6755421>

MRIS DIG - [Insurance Solutions - DIG - General - All Documents \(sharepoint.com\)](#)

MRIS DQ and OAS shared folder: [OAS Engagement](#)

UHG Guidelines Link: <https://architecture.uhg.com/docs/enterprise-technology-strategies-detailed.html#data-management>  
<https://docs.hcp.uhg.com/data-platform/edge-getting-started>

IRAAD: [Issues, Risks, Assumptions, Action Items, Decisions](#)  
Meta

**Sojan Paul**  
UnitedHealthcare Medicare & Retirement  
Insurance Solutions  
680 Blair Mill Road  
Horsham, PA 19044-2233  
215-902-8997  
[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)

**Scott D Barker**  
680 BLAIR MILL RD  
Horsham PA 19044  
USA  
[scott\\_barker\\_@uhc.com](mailto:scott_barker_@uhc.com)

AOP - Administration of Policy

# To Do Items

Monday, June 12, 2023 9:05 PM

How do we want to work - 06/12/2023

- **Create template of DQ use case - TBD**
- Create a list of things (activities) we need to do
- start with KPIs
- Complete vendor demos
- collect user rating, add scoring and ratings

## Tasks to do 06/20/2023

### Use cases

identify / List business impact of DQ - Scott to provide anecdotally - done

Claims, Member Info., Agent Info.

Tabulate data elements, process(es) impacting data elements - to be done by Kim F.

- Kim F. to provide - in-progress

Review, refine with Scott B. and Kim F.

- initial review - done
- Final review - as soon as Kim is done

Build the charter with KPI to track and implement

Data Profiling - identified fields

Key dimensions

- Data accuracy: The percentage of data that is accurate.
- Data completeness: The percentage of data that is complete and contains necessary information.
- Data consistency: The degree to which data is consistent and conforms to a set of rules and standards.
- Data timeliness: Degree to which the data is up-to-date and reflects the current state of affairs.
- Data relevance: Degree to which data is meaningful and relevant for the intended purpose.

### Tool analysis & replacement

Get demo from Talend Data Quality - done

Get second demo from Informatica - done

Score DQ tools - done

Scott review - done

Review with Dinesh and Scott separately - to be scheduled with Dinesh

Make DQ tool recommendation - ready to make recommendation

Guestimate rollout timeline

Based on the activities listed below, it will take 2-4 weeks

- Setup 1-2 weeks
- Configuration and validation - 1-2 weeks
- Choose data quality tool - done
- DQaaS and MRIS will -
  - Install data quality tool
  - Configure data quality tool
  - Run data quality checks to identify errors and inconsistencies in your data
  - Monitor data Quality to ensure DQ remains accurate and consistent
  - Get buy-in from stakeholders
  - Start small
  - Use a test dataset
  - Seek out DQaaS help to get you started

Document training option for the DQ tool

For training DQaaS will conduct trainings, initial demo and share training videos

Also, the following documentation is available:

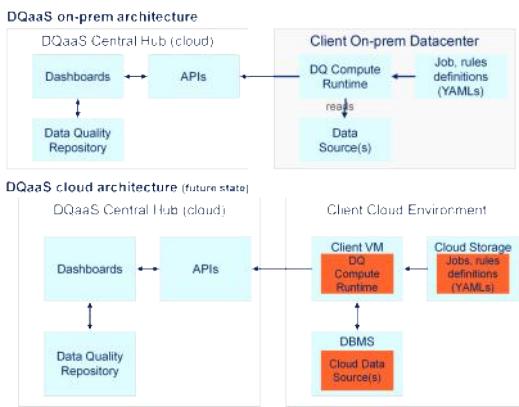
- DQaaS Product Main Page: <https://hcp.uhg.com/products/dqaas>
- Tenant Onboarding: [https://dqaas-docs.optum.com/docs/GettingStarted/onboarding\\_steps](https://dqaas-docs.optum.com/docs/GettingStarted/onboarding_steps)
- DQaaS Playbook: <https://dqaas-docs.optum.com/docs/TechnicalArticles/playbook>
- Setup DQ Jobs in console: [https://dqaas-docs.optum.com/docs/UserGuide/setup\\_DQ\\_job](https://dqaas-docs.optum.com/docs/UserGuide/setup_DQ_job)
- DQaaS Product Documentation: <https://dqaas-docs.optum.com/docs/Product/KeyFeatures>
- DQaaS Architecture: <https://dqaas-docs.optum.com/docs/Architecture/architecture>

Schedule time Gaurav, document steps for POC and with timelines and commitments-

- Intake request - done
- work scheduled to start: July 15th, 2023

Create block diagram of how DQ tool is working in two formats - on-prem and in the cloud

## DQaaS Architecture



Key components & services of DQaaS:

- DQ compute runtime:** The DQ compute runtime is a scalable and fault tolerant DQ executable that processes input rules on the configured data sets and generate the results. The DQ runtime is built to run on Kubernetes service and recommended running in tenant's environment to move the compute closer to where the data is residing. It is callable and run DQ jobs in ways such as - integrate to their data pipelines, schedule via CronJob or job orchestration tools such as Airflow or Oozie.
- Declarative YAML files (job / rule):** The YAML, declarative files, carry the tenant configuration that feeds the DQ runtime to run jobs & rules & data source to run on. These files can be created manually or generated out of the UI.
- DQ API's:** DQaaS provides API endpoints to programmatically access their DQ results. It provides mechanism to access information from and persist in central repository.
- DQ Central Repository:** A central database that stores the DQ rules, measures, aggregated scores & job details to facilitate quality analysis and reporting; but does not store actual data or column values.
- Dashboards:** DQaaS provides tailored dashboards to monitor & analyze the quality of their data sets. They are available on the console where tenants can configure jobs, rules, checks & alerts.

### Deployment model

DQaaS is designed to operate in the following deployment models:

- Near-data processing:** Move processor (DQ compute runtime) close to the data, instead of shipping the data to the processor. The DQ runtime runs on tenant's premise to connect to their storage systems.
- Software-as-a-service(SaaS) model:** The common services are made centrally available as a SaaS model for the users to subscribe and access the DQ application.
- Integration flexibility:** Tenants are provided with DQ runtime as an executable that can be integrated or called in data pipelines, from orchestration tools & other trigger mechanisms.

### SDLC

Complete and review documents with EA - scheduled for 06/28/2023 - done

Review with EA - scheduled for 06/28/2023 - done

Complete Governance KPIs - to be discussed with Anantha

Review with Governance team - scheduled for 06/28/2023 - done

Identification pilot project for DQ operationalization - to discuss with EA and Governance teams

DQ enablement in SDLC and EA - in-progress - entry points identified

Document the DQ enablement process (in SDLC and EA)

Roadmap/plan for other ongoing projects

Implement KPI cadence

\* Review with Liv of June 30th or July 5th

## Emails

Monday, May 15, 2023 11:10 AM

RE: MRIS DQ Enablement & AOP metadata Lineage - Kick off

 Barker, Scott D  
To:  Sojan K,  Ferguson, Kimberly  
Retention Policy: UHGSent (90 days)

Expires: 8/9/2023

Thu 5/11/2023 2:32 P

 Internal

Hi Sojan,

For the part about getting DQ steps into EA and SDLC process, here are some possible contacts: David Weiss, John Purcell, Mike Hastie, Dave Searfass, maybe Suwarna or Tim Weiss for picking a project or a PI planning. Do you have suggestions on who else might we want to contact? All these folks I listed are around in next 3 weeks except John in week 3.

Basically, our ask is that for new projects and at PI planning:

- Will project require data from other systems or external entities?
  - o If external entity, is B2B doc being developed? If not, need to include in plan
  - o Has data been profiled for data quality? If not, need to include in plan, as early as possible
- Will project be sending data out to other systems or external entities
  - o Has DUR document been requested?
  - o Does data to be sent include member PII data?

These are small asks, but get our foot in the door to project and PI planning.

Thanks,  
Scott

-----Original Appointment-----

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Sent:** Tuesday, May 9, 2023 10:49 AM  
**To:** Ramakrishnan, Anantha; Paul, Sojan K; Barker, Scott D; Thulasi, Rajkumar; Paramatmuni, Srilatha; Malhotra, Dinesh; Cheema, Dave; Mandal, Aditya; Ferguson, Kimberly  
**Subject:** MRIS DQ Enablement & AOP metadata Lineage - Kick off

RE: candidate data for DQ profiling

 Linga, Sujaya  
To:  Barker, Scott D;  White, Donald E;  Cookson, Timothy O;  Searfass, David R  
Retention Policy: UHGSinbox (90 days)

Expires: 8/13/2023

 Internal

 FOX\_PreBatch\_SQLs.txt  
17 KB

Start your reply all with:

Hi Scott,

Here are some sqls we run prior to overnight batch to prevent batch abends due to data issues.

Thanks,  
Sujaya

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Monday, May 15, 2023 10:39 AM  
**To:** White, Donald E <[don.white@uhc.com](mailto:don.white@uhc.com)>; Cookson, Timothy O <[timothy.o.cookson@uhc.com](mailto:timothy.o.cookson@uhc.com)>; Searfass, David R <[david.searfass@uhc.com](mailto:david.searfass@uhc.com)>  
**Cc:** Linga, Sujaya <[sujaya.linga@uhc.com](mailto:sujaya.linga@uhc.com)>  
**Subject:** RE: candidate data for DQ profiling

Hi Sujaya,

I am trying to learn about what data quality checks we may be doing today – could you send me some examples of these queries or describe what they check? If these are for FOX, I am very familiar with the tables.

\*\*\*\*\*

FW: COMPAS - multiple examples of individuals sharing the same email address

 Barker, Scott D  
To:  Cheema, Dave;  Ramakrishnan, Anantha  
Cc:  Ferguson, Kimberly;  Shekhtman, Felix  
Retention Policy: UHGSinbox (90 days)

Expires: 8/13/2023

 Internal

 COMPAS RITM3280029\_updated.docx  
24 KB

Hi Dave,

This would be our other data quality use case. Whereas Agent was straight profiling of columns, this has to do with business rules. There are valid cases where different insured members can share an email address – for example a husband and wife. But we have cases where different people who share the same name and DOB (yes, it happens) are incorrectly sharing the same E-mail. Everything else about them is different – MBI, account, household, zip code, etc.

This can have severe member impact. We have not yet discovered the process that is the root cause. We have SQL scripts to detect bad data.

Thanks,  
Scott

**From:** Barker, Scott D  
**Sent:** Wednesday, March 15, 2023 9:35 AM  
**To:** Searfass, David R <[david.searfass@uhc.com](mailto:david.searfass@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>; Owen, Susan J <[susan.owen@uhc.com](mailto:susan.owen@uhc.com)>; Paul, Sojan K <[sojan.paul@uhc.com](mailto:sojan.paul@uhc.com)>; Danner, James <[james.danner@uhc.com](mailto:james.danner@uhc.com)>; Cimino, Adam J <[adam.j.cimino@uhc.com](mailto:adam.j.cimino@uhc.com)>; McClelland, Michelle L <[micelle.l.mcclelland@uhc.com](mailto:micelle.l.mcclelland@uhc.com)>; Shekhtman, Felix <[felix.shekhtman@uhc.com](mailto:felix.shekhtman@uhc.com)>; Shah, Kunjal V <[kunjal.shah@uhc.com](mailto:kunjal.shah@uhc.com)>; Vartanian, Lauren <[lauren.vartanian@uhc.com](mailto:lauren.vartanian@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Hi Dave,

Here is a 1-page summary/justification for the COMPAS change to prevent cases where email address is being incorrectly shared among members with same name and DOB.

Thanks go to Kim!

Scott

**From:** Barker, Scott D  
**Sent:** Thursday, March 2, 2023 11:30 AM  
**To:** Searfass, David R <[david.searfass@uhc.com](mailto:david.searfass@uhc.com)>

Scott

**From:** Barker, Scott D  
**Sent:** Thursday, March 2, 2023 11:30 AM  
**To:** Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Owen, Susan J <[susan\\_owen@uhc.com](mailto:susan_owen@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Danner, James <[james\\_danner@uhc.com](mailto:james_danner@uhc.com)>; Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Shah, Kunjal V <[kunjal.shah@uhc.com](mailto:kunjal.shah@uhc.com)>; Vartanian, Lauren <[lauren.vartanian@uhc.com](mailto:lauren.vartanian@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Hi Dave,

Following up from meeting yesterday... the Service Request for COMPAS that Adam opened is RITM3280029.

This group feels that the addition of business rules in COMPAS for Individual Email Address would fix an existing problem where different individuals are incorrectly sharing an email address. While there are only around 50 cases out of 15 MM individuals, the feeling is that 1 is too many. This would improve member data quality on COMPAS and IMDM since COMPAS data flows to IMDM.

I am not sure what the intake process is for COMPAS change requests. If you could be an advocate for this change, that would be great.

Thanks,  
Scott

---

**From:** Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>  
**Sent:** Monday, February 27, 2023 3:30 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>; Owen, Susan J <[susan\\_owen@uhc.com](mailto:susan_owen@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Danner, James <[james\\_danner@uhc.com](mailto:james_danner@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Hello,

I just opened the request to add the below enhancement to COMPAS. The ticket in ServiceNow is RITM3280029.

Please let me know if you have any questions.

**Adam Cimino** | Billing Supervisor  
United Healthcare | Medicare & Retirement | Insurance Solutions  
680 Blair Mill Rd  
Horsham, PA 19044  
215-902-9944  
[Adam\\_j\\_cimino@uhc.com](mailto:Adam_j_cimino@uhc.com)

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Monday, February 27, 2023 3:04 PM  
**To:** McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>; Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>; Owen, Susan J <[susan\\_owen@uhc.com](mailto:susan_owen@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Danner, James <[james\\_danner@uhc.com](mailto:james_danner@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Hi all,

Thank you Michelle. I think with this, that Adam could go ahead and submit a Service Now ticket for COMPAS for this change.

We'll need to align on the exact business rule we want. Maybe we want to check for all 3.

For example:

- 2 different people with same names and DOB but different membership/household can't have the same email address.  
or
- 2 different people with same names and DOB but different ZIP codes can't have same email address  
Or
- 2 different people with same names and DOB but different Medicare ID's can't have same email address

Thanks,  
Scott

---

**From:** McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Sent:** Monday, February 27, 2023 6:29 AM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>; Owen, Susan J <[susan\\_owen@uhc.com](mailto:susan_owen@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Danner, James <[james\\_danner@uhc.com](mailto:james_danner@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Good morning,

It's a good idea to add a business rule to prevent members from incorrectly sharing the same email since we don't have any rules in place now to prevent this from happening. This can happen when the member contacts the call center and the email address is added to the incorrect member.

Thanks,

Michelle

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Friday, February 24, 2023 1:42 PM  
**To:** Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>; Owen, Susan J <[susan\\_owen@uhc.com](mailto:susan_owen@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Danner, James <[james\\_danner@uhc.com](mailto:james_danner@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

+ Jim Danner

Hi all,

Michelle – we wanted to get your opinion on the suggestion below about adding a business rule to COMPAS to prevent situations where 2 members with same name and DOB, but different ZIP codes and/or Medicare ID's, are incorrectly sharing the same email. As you know, we found around 50 such existing cases in

Hi all,

Michelle – we wanted to get your opinion on the suggestion below about adding a business rule to COMPAS to prevent situations where 2 members with same name and DOB, but different ZIP codes and/or Medicard ID's, are incorrectly sharing the same email. As you know, we found around 50 such existing cases in COMPAS. We wanted to check with you to see if there was any history on this, or something we are not aware of.

Dave – if the answer is yes, would next step be for Adam to submit a Service Now ticket for COMPAS?

Jim – do you know if these EIMP Over-merge cases the same as the cases MI 031 and 032 in Member Impact? If so, do you want to stay as owner, or switch to Data Governance.

Thanks,  
Scott

---

**From:** Barker, Scott D  
**Sent:** Tuesday, February 21, 2023 8:47 AM  
**To:** Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>; Owen, Susan J <[susan\\_owen@uhc.com](mailto:susan_owen@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

+ Dave and Susan,

Hi Dave,

In the discussion below, we are saying that it might be possible to add a business rule to COMPAS to prevent situations where 2 different individuals are incorrectly sharing the same email. What is proper way to initiate a request? Would it be for Adam to submit a Service Now ticket?

Thanks,  
Scott

---

**From:** Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>  
**Sent:** Tuesday, February 21, 2023 7:23 AM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Hi Scott,

I think it makes sense to pursue this effort. The parameters you have listed below should be a good start.

This would be a system enhancement for COMPAS. There wouldn't exactly be a SME, but I'm thinking it would be someone from the Development team or possibly IT Ops. We usually submit these requests through Service Now. Should I submit the request to get the process started?

**Adam Cimino** | Billing Supervisor  
United Healthcare | Medicare & Retirement | Insurance Solutions  
680 Blair Mill Rd  
Horsham, PA 19044  
215-902-9944  
[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Monday, February 20, 2023 2:10 PM  
**To:** Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Hi Adam,

We were thinking that maybe there should/could be a business rule added to COMPAS that would prevent the situation we are seeing.

For example:

- 2 different people that are not in the same membership/household can't have the same email address.  
or
- 2 different people with same names and DOB but different ZIP codes can't have same email address  
Or
- 2 different people with same names and DOB but different Medicare ID's can't have same email address

Do you think this is worth pursuing with any other SME's from COMPAS? Or could there be cases we aren't considering, or it occurs so infrequently that it is not worth making the change? We found around 100 cases in all of COMPAS. I believe the 6 or so we checked a while back with Michelle were all determined to be incorrect.

Thanks,  
Scott

---

**From:** Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>  
**Sent:** Monday, February 20, 2023 8:53 AM  
**To:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Cc:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Felix,

We weren't able to locate any rules for email addresses in COMPAS.

**Adam Cimino** | Billing Supervisor  
United Healthcare | Medicare & Retirement | Insurance Solutions  
680 Blair Mill Rd  
Horsham, PA 19044  
215-902-9944  
[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)

---

**From:** Cimino, Adam J

[Adam\\_j\\_cimino@uhc.com](mailto:Adam_j_cimino@uhc.com)

**From:** Cimino, Adam J  
**Sent:** Friday, February 10, 2023 2:49 PM  
**To:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Cc:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** RE: COMPAS - multiple examples of individuals sharing the same email address

Hi Felix,

I'm checking with our Compliance area on your question about email address rules in COMPAS. I hope to have an answer back to you early next week.

We're not aware of any rules but want to be sure before confirming yes/no.

**Adam Cimino** | Billing Supervisor  
United Healthcare | Medicare & Retirement | Insurance Solutions  
680 Blair Mill Rd  
Horsham, PA 19044  
215-902-9944  
[Adam\\_j\\_cimino@uhc.com](mailto:Adam_j_cimino@uhc.com)

---

**From:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Sent:** Thursday, February 9, 2023 3:25 PM  
**To:** Cimino, Adam J <[adam\\_j\\_cimino@uhc.com](mailto:adam_j_cimino@uhc.com)>; McClendon, Michelle L <[michelle\\_l\\_mcclendon@uhc.com](mailto:michelle_l_mcclendon@uhc.com)>  
**Cc:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** COMPAS - multiple examples of individuals sharing the same email address

Hello Adam/Michelle,

We need your assistance to review the following cases when different individuals are sharing the same email address. We had a few working sessions with you to review the report for EIMM Member Over-Merge issue, and now found more members in COMPAS (which were not part of that report).

1. We understand it is normal for different individuals to share the same email address if they are spouses or close relatives.
2. We run query in COMPAS PROD database, and there many (100+) cases when individuals from different households with the same last name, first name and DOB are sharing the same email address. What are the COMPAS rules for multiple individuals using the same email address?

Since this is production data, I can't share our findings here, but I will be able to share it when we meet.

Thanks  
Felix

+++++  
 SRA AGENT tables.pdf 94 KB

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Monday, May 15, 2023 12:39 PM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Subject:** RE: Please add new tables to ISDW and ISDW\_AOP sheet

Here is a pdf of the data model for those Agent tables.. You can see the data types are pretty relaxed... so should be interesting to profile.

Active_Base		
ID	NUMBER(10)	NOT NULL
PARTY_ID	VARCHAR2(50)	NOT NULL
PID_STATUS	VARCHAR2(25)	NOT NULL
PID_STATUS_REASON	VARCHAR2(75)	NULL
PID_BEGIN	DATE	NOT NULL
PID_END	DATE	NULL
FIRST_NAME	VARCHAR2(255)	NULL
MIDDLE_NAME	VARCHAR2(255)	NULL
LAST_NAME	VARCHAR2(255)	NULL
BUSINESS_NAME	VARCHAR2(255)	NULL
AGENT_NAME	VARCHAR2(512)	NULL
BIRTHDATE	DATE	NULL
EMPLOYEE_ID	VARCHAR2(50)	NULL
MANAGER_NAME	VARCHAR2(50)	NULL
MANAGER_PARTY_ID	VARCHAR2(255)	NULL
MGR_WID	VARCHAR2(50)	NULL
MGR_TYPE	VARCHAR2(255)	NULL
SSN	VARCHAR2(10)	NULL
TIN	VARCHAR2(10)	NULL
PARTY_TYPE	VARCHAR2(10)	NULL
CONTACT_TYPE	VARCHAR2(255)	NULL

**From:** Barker, Scott D  
**Sent:** Monday, May 15, 2023 1:15 PM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** FW: Please add new tables to ISDW and ISDW\_AOP sheet

Table definitions for the Agent tables...

Table Name	Table Definition	View Name
RTS_SUMMARY	Summary related information for the org specific transactions.	V_RTS_SUMMARY
REF_HIERARCHYTYPE	Reference data for the hierarchy type at the enterprise level	V_REF_HIERARCHYTYPE
PARTYSTATUS	This table stores the data related to party status fields such as active or inactive	V_PARTYSTATUS
EVERYTHING_BASE	Same as the Everything_Base table, but this table is only for ACA agents. Demographics for every Agent regardless of Active Status. The whole historical at agent demographics. Unsure of how far the table goes back.	V_EVERYTHING_BASE

EVERYTHING_BASE	Same as the Everything_Base table, but this table is only for ACA agents. Demographics for every Agent regardless of Active Status. The whole historical at agent demographics. Unsure of how far the table goes back.	V_EVERYTHING_BASE
ACTIVE_BASE	All Active Agent IDs and Party IDs Demographics - Both Agent and Party ID need to be active to appear in this table.	V_ACTIVE_BASE
HIERARCHY_EXT	Agent Hierarchy for Field Agents - External Agents	V_HIERARCHY_EXT
HIERARCHY_INT	Agent Hierarchy for Internal Agents - ICAs, Telesales, Group, etc. Internal agents who are UHC employees.	V_HIERARCHY_INT
PRINCIPAL_ORG	The Principal Organization Relations for the agencies we work with. Each Organization has an individual Principal who is an agent. This shows start and end dates for the Princ/Org relationships.	V_PRINCIPAL_ORG
RTS_APPT_STATE	Ready To Sell Appointments - All Licensed Appointment, Certified Agents, and the States in which they are appointed.	V_RTS_APPT_STATE
AGENT_TENURE	How long an agent has been an agent, Information about their tenure, as well as information about their sales with UHC. We don't currently use this table but the data looks like it could be very useful. We will need to research further.	V_AGENT_TENURE
AGENTTOOLKIT_DETAIL	Agent Toolkit is a resource for Agents - Personalized Branded Items to order, informational documents to provide to clients, etc. This table shows toolkit order information for agents.	V_AGENTTOOLKIT_DETAIL
AGENTTOOLKIT_SUMMARY	Toolkit Order Summary by Agent	V_AGENTTOOLKIT_SUMMARY
PARTY_COURSE_ATTEMPTS	Certification data for agents. How many times they took a particular course and how many times they passed it. We use this table only for their UHICA certification.	V_PARTY_COURSE_ATTEMPTS
ICM_RAW_A2O	Authorized to Offer Status. This is a status that gives agents certain benefits and perks based on if they have hit sales thresholds.	V_ICM_RAW_A2O
Replace IEX_EVERYTHING_BASE with VW_ProducerFunction_All AgentBusinessType	Same as the AGT.Everything_Base table, but this table is only for ACA agents. Demographics for every Agent regardless of Active Status. The whole historical at agent demographics. Unsure of how far the table goes back.	V_VW_ProducerFunction_All AgentBusinessType
AGENTTOOLKIT_USAGE	This shows criteria about last time(?) the agent logged into the Toolkit.	V_AGENTTOOLKIT_USAGE
CERT_DEAUTHORIZATION	Shows history of Agent DeAuthorization Status. Unauthorized agents are unable to sell the product. Status may be DeAuthed or ReAuthed.	V_CERT_DEAUTHORIZATION
LICENSE	Historical Agent Licensing Information. Start and End dates of licenses, the State it applies to, etc.	V_LICENSE
WV_DART_SMRTSALES_SOT	*This is a view within the SRA database that feeds the SMRT Sales Instance. Our team accesses the SMRT Sales through a UI called Qlikview	V_WV_DART_SMRTSALES_SOT

- Also, Kindly make the below change to the below table and views on the ISDW and ISDW\_AOP sheet:

Schema	SRA	
Roles:	R_G_OPTUM_ANAL R_G_DATA_ANAL	
Table Name	Table Definition	View Name
Replace IEX_EVERYTHING_BASE with VW_ProducerFunction_All AgentBusinessType	Same as the AGT.Everything_Base table, but this table is only for ACA agents. Demographics for every Agent regardless of Active Status. The whole historical at agent demographics. Unsure of how far the table goes back.	V_VW_ProducerFunction_All AgentBusinessType
Add New Table and View: SalesGeo	This table holds Geographical Market and Region information. Can be used to tie State, Region and Market Identifiers to Sales.	V_SalesGeo
Add New Table and View: CSR	This table includes Agent State Appointment and Licensing Information. It tracks agents who are 'Compliant_To_Sell' which tracks agent that either meeting the Ready to Sell rules OR the Just In Time appointment rules. This is a more encompassing table for tracking agents who are able to sell by state.	V_CSR

Thanks and Regards,  
Hari.

**From:** Bey, Roxanne A <[roxanne\\_bey@uhc.com](mailto:roxanne_bey@uhc.com)>  
**Sent:** Wednesday, April 26, 2023 2:25 PM  
**To:** Karnam, Hari <[hari\\_karnam@uhc.com](mailto:hari_karnam@uhc.com)>  
**Cc:** Thulasi, Rajkumar <[rajkumar\\_thulasi@uhc.com](mailto:rajkumar_thulasi@uhc.com)>  
**Subject:** RE: Please add new tables to ISDW and ISDW\_AOP sheet

Hi Hari,  
The tables identified below were added as requested.  
Roxie

**From:** Karnam, Hari <[hari\\_karnam@uhc.com](mailto:hari_karnam@uhc.com)>  
**Sent:** Wednesday, April 26, 2023 9:52 AM  
**To:** Bey, Roxanne A <[roxanne\\_bey@uhc.com](mailto:roxanne_bey@uhc.com)>

Roxie

**From:** Karnam, Hari <[hari\\_karnam@uhc.com](mailto:hari_karnam@uhc.com)>  
**Sent:** Wednesday, April 26, 2023 9:52 AM  
**To:** Bey, Roxanne A <[roxanne\\_bey@uhc.com](mailto:roxanne_bey@uhc.com)>  
**Cc:** Thulasi, Rajkumar <[rajkumar\\_thulasi@uhc.com](mailto:rajkumar_thulasi@uhc.com)>  
**Subject:** Please add new tables to ISDW and ISDW\_AOP sheet

Good morning, Roxy.

Please add the below tables, views and roles to the ISDW and ISDW AOP sheet.

Schema	ETL_FF	
Roles:	R_G_OPTUM_ANAL R_G_DATA_ANAL	
Table Name	Table Definition	View Name
RTS_SUMMARY	Summary related information for the org specific transactions.	V_RTS_SUMMARY
REF_HIERARCHYTYPE	Reference data for the hierarchy type at the enterprise level	V_REF_HIERARCHYTYPE
PARTYSTATUS	This table stores the data related to party status fields such as active or inactive	V_PARTYSTATUS
EVERYTHING_BASE	Same as the Everything_Base table, but this table is only for ACA agents. Demographics for every Agent regardless of Active Status. The whole historical at agent demographics. Unsure of how far the table goes back.	V_EVERYTHING_BASE
ACTIVE_BASE	All Active Agent IDs and Party IDs Demographics - Both Agent and Party ID need to be active to appear in this table.	V_ACTIVE_BASE
HIERARCHY_EXT	Agent Hierarchy for Field Agents - External Agents	V_HIERARCHY_EXT
HIERARCHY_INT	Agent Hierarchy for Internal Agents - ICAs, Telesales, Group, etc. Internal agents who are UHC employees.	V_HIERARCHY_INT
PRINCIPAL_ORG	The Principal Organization Relations for the agencies we work with. Each Organization has an individual Principal who is an agent. This shows start and end dates for the Princ/Org relationships.	V_PRINCIPAL_ORG
RTS_APPT_STATE	Ready To Sell Appointments - All Licensed Appointment, Certified Agents, and the States in which they are appointed.	V_RTS_APPT_STATE
AGENT_TENURE	How long an agent has been an agent, Information about their tenure, as well as information about their sales with UHC. We don't currently use this table but the data looks like it could be very useful. We will need to research further.	V_AGENT_TENURE
AGENTTOOLKIT_DETAIL	Agent Toolkit is a resource for Agents - Personalized Branded Items to order, informational documents to provide to clients, etc. This table shows toolkit order information for agents.	V_AGENTTOOLKIT_DETAIL
AGENTTOOLKIT_SUMMARY	Toolkit Order Summary by Agent	V_AGENTTOOLKIT_SUMMARY
PARTY_COURSE_ATTEMPTS	Certification data for agents. How many times they took a particular course and how many times they passed it. We use this table only for their UHICA certification.	V_PARTY_COURSE_ATTEMPTS
ICM_RAW_A2O	Authorized to Offer Status. This is a status that gives agents certain benefits and perks based on if they have hit sales thresholds.	V_ICM_RAW_A2O
Replace IEX_EVERYTHING_BASE with VW_ProducerFunction_All AgentBusinessType	Same as the AGT.Everything_Base table, but this table is only for ACA agents. Demographics for every Agent regardless of Active Status. The whole historical at agent demographics. Unsure of how far the table goes back.	V_VW_ProducerFunction_All AgentBusinessType
AGENTTOOLKIT_USAGE	This shows criteria about last time(?) the agent logged into the Toolkit.	V_AGENTTOOLKIT_USAGE
CERT_DEAUTHORIZATION	Shows history of Agent DeAuthorization Status. Unauthorized agents are unable to sell the product. Status may be Deauthed or ReAuthed.	V_CERT_DEAUTHORIZATION
LICENSE	Historical Agent Licensing Information. Start and End dates of licenses, the State it applies to, etc.	V_LICENSE
WV_DART_SMRTSALES_SOT	*This is a view within the SRA database that feeds the SMRT Sales Instance. Our team accesses the SMRT Sales through a UI called Qlikview	V_WV_DART_SMRTSALES_SOT

Thanks and Regards,  
Hari

\*\*\*\*\*

#### FW: SRA Application - Question on Agent PII

 Barker, Scott D  
To: Ramakrishnan, Anantha; Cheema, Dave; Mandal, Aditya  
Cc: Ferguson, Kimberly

Retention Policy: UHGlinbox (90 days)

Expires: 8/13/2023

Internal

You replied to this message on 5/15/2023 12:01 PM.

Mon 5/15/2023 11:55 AM

Hi Dave,

Agent is looking like a good use case for the DQ tool evaluation and process work. The volumes are manageable, no member PII, and data quality can vary.

We have a copy in our data warehouse development area. We also have table definitions which I will forward.

Scott

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>

From: Barker, Scott D <scott\_barker@uhc.com>

Scott

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Friday, May 12, 2023 1:55 PM  
**To:** Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Thanks Arjunan.

I think these SRA Agent tables will fit the bill for data quality profiling and testing. Reasonable volumes, no member PII, will be new source to ISDW, we have table definitions.

Is this data anywhere where Felix could look at it?

]

Arjunan – just fyi... for profiling we will want the real data – not masked.

Thanks,  
Scott

**From:** Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>  
**Sent:** Friday, May 12, 2023 10:13 AM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Hi Scott,

- Are these all new source tables for ISDW? Yes
- Do we have a set of populated tables? Yes, for our unit testing, but we would have to wait for masking solution to be implemented before using those tables. The ETL code would be in production May 23<sup>rd</sup>, so these new tables will have data in PROD.
- Is volume of data in these tables small – like as compared to claims? Volume is very less, below I have given the count from SRA production, its less than a million for most of the tables and few millions for couple of tables.

SNO	Database	Tables	ROWS
1	SRA	AGT.RTS_SUMMARY	883856
2	SRA	AGT.PARTYSTATUS	1674796
3	SRA	AGT.EVERYTHING_BASE	644687
4	SRA	AGT.ACTIVE_BASE	111788
5	SRA	AGT.HIERARCHY_EXT	673796
6	SRA	AGT.HIERARCHY_INT	60526
7	SRA	AGT.PRINCIPAL_ORG	27710
8	SRA	AGT.RTS_APPT_STATE	670292
9	SRA	AGT.AGENT_TENURE	458613
10	SRA	AGT.AGENTTOOLKIT_DETAIL	1736918
11	SRA	AGT.AGENTTOOLKIT_SUMMARY	1602559
12	SRA	AGT.PARTY_COOURSE_ATTEMPTS	6634362
13	SRA	ICM.Raw_A2O	662953
14	SRA	IEX.Everything_Base	167410
15	SRA	APP.VW_DART_SMRTSALES_SOT	13062129
16	SRA	AGT.REF_HIERARCHYTYPE	11
17	SRA	AGT.AGENTTOOLKIT_USAGE	370189
18	SRA	AGT.CERT_DEAUTHORIZATION	33968
19	SRA	AGT.LICENSE	6775185
20	SRA	AGT.VM_PRODUCERFUNCTION_ALL	561393
21	SRA	AGT.CSR	1644106
22	SRA	APP.SalesGeo	51524
23	SRA	AGT.ProdInfoType	643143

Thank you,  
Arjunan

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Friday, May 12, 2023 9:47 AM  
**To:** Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Hi Arjunan,

We are looking for a subject area for a possible data quality POC, and I was wondering if these Agent tables would be good candidates. Couple questions:

- Are these all new source tables for ISDW?
- Do we have a set of populated tables?
- Is volume of data in these tables small – like as compared to claims?

Thanks,  
Scott

**From:** Barker, Scott D  
**Sent:** Thursday, May 4, 2023 9:21 AM  
**To:** Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>; Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>; Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>  
**Cc:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>; Maggs, Robert E <[robert.e.maggs@uhc.com](mailto:robert.e.maggs@uhc.com)>; Nallam, Nageswara Rao <[nages.nallam@uhc.com](mailto:nages.nallam@uhc.com)>; Koreka, Nishith <[nishith.koreka@uhc.com](mailto:nishith.koreka@uhc.com)>; Tarapareddy, Vamsi Krishna <[vamsi.krishna.tarapareddy@uhc.com](mailto:vamsi.krishna.tarapareddy@uhc.com)>; Mathai, Jessy <[jessy.mathai@uhc.com](mailto:jessy.mathai@uhc.com)>; Raveendiran, Vinoth Kumar <[vinoth.raveendiran@uhc.com](mailto:vinoth.raveendiran@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Hi Arjunan,

A couple of these appear to be reference tables – like SALESgeo and PRODINFOType – and I think are ok.

The main one I see is VW\_DART\_SMRTSALES\_SOT. I marked the PII columns, but if this is a view pulling from other base tables, we should have those columns already marked/masked in base tables.

AGENTTOOLKIT_USAGE	Looks fine - Reference table
CERT_DEAUTHORIZATION	Has a party name

The main one I see is **VW\_DART\_SMRTSALES\_SOT**. I marked the PII columns, but if this is a view pulling from other base tables, we should have those columns already marked/masked in base tables.

AGENTTOOLKIT_USAGE	Looks fine - Reference table
CERT_DEAUTHORIZATION	Has a party name
CSR	Looks fine
LICENSE	Has a license number
PRODINFOTYPE	Looks fine - Reference table
SALESGEO	Looks fine - reference table
VW_DART_SMRTSALES_SOT	Is a view, has multiple PII elements
VW_PRODUCERFUNCTION_ALL	Is a view, has a partner name, need definition of FMONAME. Is an FMO a person?

Thanks,  
Scott

---

**From:** Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>  
**Sent:** Wednesday, May 3, 2023 2:33 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>; Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>  
**Cc:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>; Maggs, Robert E <[robert.e.maggs@uhc.com](mailto:robert.e.maggs@uhc.com)>; Nallam, Nageswara Rao <[nages.nallam@uhc.com](mailto:nages.nallam@uhc.com)>; Koreka, Nishith <[nishith.koreka@uhc.com](mailto:nishith.koreka@uhc.com)>; Tarapareddy, Vamsi Krishna <[vamsi.krishna.tarapareddy@uhc.com](mailto:vamsi.krishna.tarapareddy@uhc.com)>; Mathai, Jessy <[jessy.mathai@uhc.com](mailto:jessy.mathai@uhc.com)>; Raveendiran, Vinoth Kumar <[vinoth.raveendiran@uhc.com](mailto:vinoth.raveendiran@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Hi Scott,

Business have added eight more tables, I have included them in a new sheet (Table Columns - List2). Could you please check and share your suggestion on masking.

Thank you,  
Arjunan

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Tuesday, May 2, 2023 11:57 AM  
**To:** Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>; Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>  
**Cc:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>; Maggs, Robert E <[robert.e.maggs@uhc.com](mailto:robert.e.maggs@uhc.com)>; Nallam, Nageswara Rao <[nages.nallam@uhc.com](mailto:nages.nallam@uhc.com)>; Koreka, Nishith <[nishith.koreka@uhc.com](mailto:nishith.koreka@uhc.com)>; Tarapareddy, Vamsi Krishna <[vamsi.krishna.tarapareddy@uhc.com](mailto:vamsi.krishna.tarapareddy@uhc.com)>; Mathai, Jessy <[jessy.mathai@uhc.com](mailto:jessy.mathai@uhc.com)>; Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Hi all,

I have marked all the Agent PII columns in the attached spreadsheet - if/since the direction is that all Agent PII data is to be masked.

I believe last week Jessy indicated that if this data needs masking, then a request needs to come to masking team through normal intake channel.

Thanks,  
Scott

---

**From:** Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>  
**Sent:** Tuesday, May 2, 2023 11:18 AM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>  
**Cc:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>; Maggs, Robert E <[robert.e.maggs@uhc.com](mailto:robert.e.maggs@uhc.com)>; Nallam, Nageswara Rao <[nages.nallam@uhc.com](mailto:nages.nallam@uhc.com)>; Koreka, Nishith <[nishith.koreka@uhc.com](mailto:nishith.koreka@uhc.com)>; Tarapareddy, Vamsi Krishna <[vamsi.krishna.tarapareddy@uhc.com](mailto:vamsi.krishna.tarapareddy@uhc.com)>; Mathai, Jessy <[jessy.mathai@uhc.com](mailto:jessy.mathai@uhc.com)>; Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Hi Scott and Preetha,  
All Agent data in either the Test or Production environment should be masked as "Protected" and "Proprietary". Proprietary is labeled as "Confidential" on the ISDW mega mapping document. Please confirm this will be the approach.  
Thank you for your patience while this was being reviewed.

Rose

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Tuesday, May 2, 2023 8:47 AM  
**To:** Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>; Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>  
**Cc:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>; Maggs, Robert E <[robert.e.maggs@uhc.com](mailto:robert.e.maggs@uhc.com)>; Nallam, Nageswara Rao <[nages.nallam@uhc.com](mailto:nages.nallam@uhc.com)>; Koreka, Nishith <[nishith.koreka@uhc.com](mailto:nishith.koreka@uhc.com)>; Tarapareddy, Vamsi Krishna <[vamsi.krishna.tarapareddy@uhc.com](mailto:vamsi.krishna.tarapareddy@uhc.com)>; Mathai, Jessy <[jessy.mathai@uhc.com](mailto:jessy.mathai@uhc.com)>; Easwaran, Arjunan <[arjunan.easwaran@uhc.com](mailto:arjunan.easwaran@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

+ Arjunan

Hi all,

As I mentioned on separate thread, a group of us had met about this last week. In summary we thought that since this was Agent data and not Member data, then masking was not needed. But this information below would override that. My understanding is that Ally is saying there is agent business relationship (commission structures), and that is what warrants the masking.

Two points of clarification, I believe these are 15 new tables – different from the 4 tables that Rose mentions below. And this is for masking in lower environments, not production access controlled by the Secure roles.



Thanks,  
Scott

---

**From:** Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>  
**Sent:** Monday, May 1, 2023 5:15 PM  
**To:** Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>; Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>

Scott

**From:** Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>  
**Sent:** Monday, May 1, 2023 5:15 PM  
**To:** Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>; Barker, Scott D <[scott.barker@uhc.com](mailto:scott.barker@uhc.com)>  
**Cc:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>; Maggs, Robert E <[robert.e.maggs@uhc.com](mailto:robert.e.maggs@uhc.com)>; Nallam, Nageswara Rao <[nages.nallam@uhc.com](mailto:nages.nallam@uhc.com)>; Koreka, Nishith <[nishith.koreka@uhc.com](mailto:nishith.koreka@uhc.com)>; Tarapareddy, Vamsi Krishna <[vamsi.krishna.tarapareddy@uhc.com](mailto:vamsi.krishna.tarapareddy@uhc.com)>; Mathai, Jessy <[jessy.mathai@uhc.com](mailto:jessy.mathai@uhc.com)>  
**Subject:** RE: SRA Application - Question on Agent PII

Hi Preetha & Scott,  
Ally forwarded your request to Rob and I on whether the non-prod Agent data should be masked. I researched the HIPAA and Privacy regulations on CMS, HHS, Safe Harbor and HITECH. It took a little more research than anticipated as we were unclear as to what the SRA application was and whether it included member data on the same file. After reviewing the details from your Team, we found SRA is an ISDW schema solely for Agent data. I shared the field level detail and classifications with Ally. After reviewing the Agent information, Ally advised Masking all of the Non-Prod Agent data and changing the ISDW SRA schema's "table classification" to "Protected" and the "contains protected PHI/PII" to "Confidential" since it includes proprietary Agent contract info.  
Ally's feedback,

"I wouldn't consider it PHI, but I'd consider it protected information – possibly even proprietary"

This leads us to recommending masking ALL of the Agent Non-Prod data as it includes proprietary data. The ISDW Schema "SRA" currently includes 4 Tables and 69 columns.

- 4 Tables:
  - ContactPrimary;
  - Party;
  - VW\_Hierarchy\_Names
  - Writing
- 69 Columns

A little more detail:

- 4 UHG Protected fields listed: Last Name, Middle Name, First Name and DOB
- 5 Secure Roles with access: ISDW Actuarial Healthcare Economics, ISDDW Data Analytics, ISDW Finance - Analytics, ISDW Agent / eAlliance, and ISDW Marketing Technology

Ally indicated a meeting can be scheduled if we need to clarify anything. I'm cc'ing Scott as this involves the ISDW Mega Mapping and we just discussed this recommendation. Scott will share his updates. Please let me know if you have any questions and if needed we'll schedule a meeting. Thank you.

Rose

**From:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>  
**Sent:** Tuesday, April 25, 2023 4:21 PM  
**To:** Corr, Rosemary R <[rosemary.corr@uhc.com](mailto:rosemary.corr@uhc.com)>; Maggs, Robert E <[robert.e.maggs@uhc.com](mailto:robert.e.maggs@uhc.com)>  
**Subject:** FW: SRA Application - Question on Agent PII

**From:** Rao, Preetha <[preetha.rao@uhc.com](mailto:preetha.rao@uhc.com)>  
**Sent:** Tuesday, April 25, 2023 2:57 PM  
**To:** Haley, Allison <[ally.haley@uhc.com](mailto:ally.haley@uhc.com)>  
**Cc:** Nallam, Nageswara Rao <[nages.nallam@uhc.com](mailto:nages.nallam@uhc.com)>; Koreka, Nishith <[nishith.koreka@uhc.com](mailto:nishith.koreka@uhc.com)>; Tarapareddy, Vamsi Krishna <[vamsi.krishna.tarapareddy@uhc.com](mailto:vamsi.krishna.tarapareddy@uhc.com)>; Mathai, Jessy <[jessy.mathai@uhc.com](mailto:jessy.mathai@uhc.com)>  
**Subject:** SRA Application - Question on Agent PII

Hi Allison,  
We are part of Test Data Management team and masking is one of our main areas.  
We are working with SRA application on masking their data in non production environment.

SRA application contains Agent's PII some of which are his Full Name, DOB , ADDRESS , PHONE , EMAIL  
Can you help us with confirming if Agent information comes under compliance and need to be masked if stored in a Non-Prod Environment.

Thanks

```
PID_END
FIRST_NAME
MIDDLE_NAME
LAST_NAME
BUSINESS_NAME
AGENT_NAME
BIRTHDATE
EMPLOYEE_ID
MANAGER_NAME
MANAGER_PARTY_ID
MGR_WID
MGR_TYPE
SSN
TIN
PARTY_TYPE
CONTACT_TYPE
ADDRESS1
ADDRESS2
+++++
```

**From:** Barker, Scott D <[scott.barker@uhc.com](mailto:scott.barker@uhc.com)>  
**Sent:** Monday, May 22, 2023 7:16 AM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>; Shekhtman, Felix <[felix.shekhtman@uhc.com](mailto:felix.shekhtman@uhc.com)>  
**Subject:** RE: Please add new tables to ISDW and ISDW\_AOP sheet

Hi Dave – as far as I know, this is a COTS package we use.

Scott

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Sent:** Saturday, May 20, 2023 7:03 PM  
**To:** Barker, Scott D <[scott.barker@uhc.com](mailto:scott.barker@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>; Shekhtman, Felix <[felix.shekhtman@uhc.com](mailto:felix.shekhtman@uhc.com)>  
**Subject:** RE: Please add new tables to ISDW and ISDW\_AOP sheet

Scott,

Please educate me. I see the following discrepancies in the state column in various tables shown below. Please help me understand the rationale of having these discrepancies:

Table name(s)	Column name	Data type
IEX_Everything_Base, Everything_Base, Active_Base	STATE	varchar(20)
CSR	STATE	varchar(50)
SalesGeo, AgentToolKit_Detail	STATE	varchar(2)
SalesGeo	StateZipCode	varchar(10)
...	...	...

IC_A_Everything_Base, Everything_Base, Active_Base	STATE	varchar(20)
CSR	STATE	varchar(50)
SalesGeo, AgentToolKit_Detail	STATE	varchar(2)
SalesGeo	StateZipCode	varchar(10)
license	statename residentstate	varchar(50)

Also, is there a State Code reference table somewhere? Thanks in advance.

Regards,

Dave Cheema

(952) 205-0802

---

#### RE: MRIS DQ project items that we would need to cover



Barker, Scott D

To: Cheema, Dave; Ferguson, Kimberly  
Cc: Ramakrishnan, Anantha; Mandal, Aditya

Retention Policy: UHGlinbox (90 days)

Expires: 8/22/2023

Internal

Start your reply all with: [Thanks for the feedback!](#) [This is great, thank you!](#) [See comments below.](#) [Feedback](#)

[Reply](#) [Reply All](#) [Forward](#) [UHC](#) [...](#)

Wed 5/24/2023 1:21 PM

Reviewing now so figured I'd write my thoughts down...

- For any tools that would be priced based on "seats", we would want to estimate 5 seats or so. We think tool would be mostly used just by us technicians
- I personally am skeptical about the Machine Learning in this context. A lot has to be put in place to make that fly.
- I would put dashboard functionality as low on list
- Highest priorities for me are 1, 3, 4 and 7
- 2 and 6 are wildcards for me – I can explain

Thanks,  
Scott

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Sent:** Tuesday, May 23, 2023 6:35 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Subject:** RE: MRIS DQ project items that we would need to cover

Scott,

I had collected similar information as well and was waiting for Anantha review before I could share it with anybody. Anyhow, here it is::

#### Data Quality Tools Evaluation Criteria:

1. Data Profiling and Assessment: The tool should offer comprehensive data profiling capabilities to assess the quality and characteristics of data. It should identify data anomalies, such as duplicates, inconsistencies, and missing values, and provide data quality metrics.
2. Data Cleansing and Standardization: Look for tools that provide advanced data cleansing and standardization capabilities. The tool should support a wide range of data transformation functions, including parsing, formatting, validation, and enrichment.
3. Data Integration and Connectivity: Consider the tool's ability to integrate with various data sources, systems, and platforms. It should support multiple data integration methods, such as batch processing, real-time streaming, and API connectivity.
4. Data Quality Rules and Monitoring: The tool should allow users to define and implement data quality rules and workflows. It should offer monitoring and alerting capabilities to proactively identify and address data quality issues in real-time.
5. Data Quality Dashboards and Reporting: Evaluate the tool's reporting and visualization capabilities. It should provide intuitive dashboards and reports to track data quality metrics, trends, and issues. Look for customizable reporting options to meet specific business requirements.
6. Data Governance and Metadata Management: Consider whether the tool supports data governance practices and metadata management. It should provide features for data lineage, data stewardship, data cataloging, and data privacy compliance.
7. Scalability and Performance: Assess the tool's ability to handle large volumes of data and its performance in terms of processing speed and scalability. It should be capable of processing data at scale without compromising performance.
8. Machine Learning and Automation: Look for tools that leverage machine learning and automation techniques to improve data quality processes. These features can assist in identifying patterns, suggesting data quality rules, and automating data cleansing tasks.
9. Ease of Use and User Interface: Evaluate the tool's user interface and ease of use. It should have an intuitive design that enables both technical and non-technical users to navigate and utilize the tool effectively.
10. Vendor Viability and Support: Consider the vendor's reputation, financial stability, and customer support services. Evaluate their track record in delivering timely updates, addressing customer issues, and providing comprehensive documentation and training.
11. Cost and ROI: Finally, consider the total cost of ownership, including licensing fees, implementation costs, and ongoing maintenance. Assess the tool's potential return on investment (ROI) by considering the value it can provide in terms of improved data quality, reduced errors, and increased efficiency.

Regards,

Dave Cheema

(952) 205-0802

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Tuesday, May 23, 2023 2:54 PM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Subject:** RE: MRIS DQ project items that we would need to cover

I had to give my name, but was able to download Gartner's review last year.

Gartner research on this market focuses on technologies and approaches delivering on the current and future needs of end users. Gartner defines the data quality solutions market as stand-alone software products that address the following data quality capabilities that organizations need in their data management solutions portfolio:

- Connectivity: Access and apply data quality across a wide range of data sources, including internal/external, at rest/streaming, on-premises/cloud, relational/nonrelational data sources.
- Profiling: The statistical analysis of diverse datasets including structured/unstructured and on-premises/cloud datasets to give business users insight into the quality of data and enable them to identify data quality issues.
- Analytics and interactive visualization: Interactive analytical workflow and visual output of statistical analysis to help business and IT users identify, understand and monitor data quality issues and discover patterns and trends over time, such as through reports, scorecards, dashboards and mobile devices.
- Monitoring and detection: Perform data quality monitoring based on preconfigured, custom-built monitoring rules, or adaptive rules, and alert violations. Automatically detect outliers, anomalies, patterns and drifts. Also provide monitoring dashboard, log

- Analytics and interactive visualization: Interactive analytical workflow and visual output of statistical analysis to help business and IT users identify, understand and monitor data quality issues and discover patterns and trends over time, such as through reports, scorecards, dashboards and mobile devices.
- Monitoring and detection: Perform data quality monitoring based on preconfigured, custom-built monitoring rules, or adaptive rules, and alert violations. Automatically detect outliers, anomalies, patterns and drifts. Also provide monitoring dashboard, log files or audit trail for compliance requirements.
- Parsing, standardizing and cleansing: The decomposition and formatting of diverse datasets based on government, industry or local standards, business rules, knowledge bases, metadata, and machine learning (ML). Modification of data values to comply with domain restrictions, integrity constraints or other business rules.
- Matching, linking and merging: Match, link and merge related data entries within or across diverse datasets using a variety of traditional and new approaches such as rules, algorithms, metadata, artificial intelligence (AI) and ML.
- Metadata and lineage: Collect and discover metadata or import metadata from third-party tools. Build or import lineage to perform rapid root cause analysis of data quality issues and impact analysis of remediation. Apply active metadata findings, make use of metadata-based rule recommendations and associations, data discovery and cataloging, metrics view against critical data elements.
- Multidomain support: Address multiple data subject areas (such as various master data domains and vertical industry domains) and offer depth of packaged support (such as prebuilt data quality rules) for these subject areas.
- Address validation/geocoding: Support location-related data standardization and cleansing. Completion for partial data in real-time or batch process.
- Data curation and enrichment: Integrate externally sourced data to improve completeness and add value.
- Rule management and data validation: Design, create and deploy business rules for specific data values. The rules can be called within the solution or by third-party applications for data validation purposes, which can be done in batch or real-time mode.
- Business-driven workflow and issue resolution: The processes and user interface to manage the data quality issue resolution through the stewardship workflow. Easily identify, quarantine, assign, escalate and resolve data quality issues, facilitated by collaboration, pervasive monitoring and case management.
- DataOps support: Collaboration of data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization.
- Deployment environment: Styles of deployment, hardware, operating system and maintenance options for deploying data quality operations.
- Architecture and integration: Commonality, consistency, and interoperability among various components of the data quality toolset (including third-party tools) and other data management solutions or components.
- Automation and augmentation: Automate data quality processes or perform in-depth and rapid execution of data quality actions on datasets by leveraging advanced technologies such as AI/ML, knowledge graph, active metadata or NLP to minimize manual effort.
- Multipersona usability: Suitability of the solution to engage and support the various roles (especially nontechnical business roles) required in a data quality initiative: data engineers, stewards, DQ analysts, data architects, DI analysts, business analysts, data preparation and so on.

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Sent:** Monday, May 22, 2023 6:33 PM

**To:** Barker, Scott D <[scott\\_barker @uhc.com](mailto:scott_barker @uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>

**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>

**Subject:** RE: MRIS DQ project items that we would need to cover

Scott,

In addition to the item we would need to cover, we would also need your help to provide the weightage for the product features for the DQ tool.

Please see the table below:

Item #	Product Assessment Framework	Weightage
6	<b>Product Features and Functionality</b>	
1.2	Evaluate the core features and functionality of the product	
1.3	Determine how well the product performs its intended purpose	
1.3	Assess any unique or innovative features that differentiate it from competitors	
2	<b>User Experience (UX) and Design</b>	
2.1	Evaluate the overall user experience of the product	
2.2	Consider factors such as ease of use, intuitiveness, and accessibility	
2.3	Assess the visual design, aesthetics, and overall user interface	
3	<b>Technical Assessment</b>	
3.1	Evaluate the technical aspects of the product, such as scalability, reliability, and security	
3.2	Consider the underlying technology stack and infrastructure	
3.3	Assess any potential limitations or challenges related to implementation or maintenance	
4	<b>Value Proposition</b>	
4.1	Define the unique value the product offers to customers.	
4.2	Assess how the product solves a specific problem or meets a need.	
4.3	Identify the key benefits and advantages of the product over existing alternatives.	
5	<b>Financial Analysis</b>	
5.1	Evaluate the financial viability of the product.	
5.2	Assess revenue projections, cost estimates, and return on investment.	
5.3	Consider factors such as break-even point and payback period.	
6	<b>Risk Assessment</b>	
6.1	Identify potential risks and challenges associated with the product	
6.2	Assess factors such as market competition, regulatory compliance, and intellectual property.	
6.2	<u>Define the risk mitigation plan</u>	
7	<b>Feedback and Validation</b>	
7.1	Gather feedback from potential customers or target audience.	
7.2	Conduct user testing and prototype validation.	
7.3	Incorporate feedback into the assessment process.	
8	<b>Benefits of using a product assessment framework</b>	
8.1	It can help you to identify the strengths and weaknesses of your product.	
8.2	It can help you to understand your target market and their needs.	
8.3	It can help you to identify your competitors and their strengths and weaknesses.	
8.4	It can help you to develop a marketing and sales strategy that will reach your target market.	
8.5	It can help you to measure the success of your product and make necessary adjustments.	

Thank you.

Regards,

Dave Cheema

(952) 205-0802

**From:** Cheema, Dave

**Sent:** Monday, May 22, 2023 12:16 PM

**To:** Barker, Scott D <[scott\\_barker @uhc.com](mailto:scott_barker @uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>

**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>

**From:** Cheema, Dave  
**Sent:** Monday, May 22, 2023 12:16 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Subject:** MRIS DQ project items that we would need to cover

Scott,

Please see items below that we would need to cover for our MRIS DQ project:

Use case Analysis		
Identify all the current DQ use cases (agent, complex COMPAS email)	M/O	2
Classify the reactive and proactive use cases	M/O	2
identify / List business impact of DQ	M/O	2
Tabulate data elements, process that impact the elements	O	
identify data / Process owners (will ask around)	M/O	1
Data model analysis / Walkthrough		
	O	2
Tool Analysis		
Identify top 3 tools of choice for DQ	M/O	1
Current TCO for DQ / Metadata tool	M	1
Operating model, R & R	M	1
Projected demand	M	1
Data migration needs / volume / Set up estimation	M	1
Automation consideration	O /M	
Shortlist data lineage business Use case (4 ) for AOP		
	M /O	2
Tool for metadata finalization, Template finalization review		
	M /O	1
SDLC Process		
Analysis of development process	O /M	2
Current gaps for DQ / DIG in development process	O/M	2
Analyze Architecture process	O /M	1
Gap identification for DQ / DIG in Architecture process	O /M	1
Identify the beta projects for DQ operationalization (2 or 3)	M	1
Review meetings		
Identification and documentation of all source / target data elements for identified use cases		
	M /O	2
Sample lineage review		
	M	1

Regards,

Dave Cheema

(952) 205-0802

\*\*\*\*\*  
RE: Question on new data sources requested for ISDW

 Shekhtman, Felix  
To: Ramakrishnan, Anantha; Barker, Scott D; Cheema, Dave; Ferguson, Kimberly; Paul, Sojan K

Retention Policy: UHGLinbox (90 days)

Expires: 8/23/2023

Thu 5/25/2023 3:23 PM

 Internal

You forwarded this message on 5/26/2023 9:33 AM.

Sure Ananth.

During my QA time, I frequently checked these columns - account number, claim number, first name, last name, address line 1, city, state, zip code, amounts (paid, benefit, deductible, etc.), dates (service date, paid date, etc.).

Thanks  
Felix

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>

**Sent:** Thursday, May 25, 2023 3:13 PM

**To:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan.paul@uhc.com](mailto:sojan.paul@uhc.com)>

**Subject:** RE: Question on new data sources requested for ISDW

Hi Felix

Thank you. Could you please share a list of fields that are more frequently checked ?, In other words, items susceptible for quality challenges ?

Regards  
Ananth

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum  
M 1-551-358-4609  
[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

**Optum**

**From:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>

## Optum

**From:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Sent:** Thursday, May 25, 2023 10:44 AM  
**To:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>  
**Subject:** RE: Question on new data sources requested for ISDW

Hi Ananth,

When the data is brought from source to target – it would be placed to the Staging area as is. When ETL process takes it from the Staging area and load them to the final destinations – that is when we (as QA) check integrity for some columns like Names (First and Last), Account Number, Address lines, etc.

Thanks  
Felix

---

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Sent:** Thursday, May 25, 2023 10:27 AM  
**To:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>  
**Subject:** RE: Question on new data sources requested for ISDW

Felix , Scott, Sojan – Thank you

Felix along with this check are there any processes that check the integrity of some specific fields like , blanks, special char etc ?.

Regards  
**Ananth**

---

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum

M 1-551-358-4609  
[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

## Optum

**From:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Sent:** Thursday, May 25, 2023 9:15 AM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>  
**Subject:** RE: Question on new data sources requested for ISDW

Hi Scott,

I am not aware of any practice to profile new data sources before bringing them to the target database. From my QA experience – we performed a complete testing after completion of ETL (bulk and incremental loads), comparing source to the target. That would include running queries in both source and target for counts and data comparison.

Thanks  
Felix

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Thursday, May 25, 2023 8:51 AM  
**To:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Subject:** RE: Question on new data sources requested for ISDW

Hi Sojan,

My experience in DW is that data profiling of new sources is a best practice and an absolute must. I recall many times bringing in new datasets, only to find that important fields were blank, or sparse, or not what user expected.

From your time in ISDW-land, do you know if data quality checking new candidate sources is part of their process for onboarding new sources?

Felix – do you happen to know?

Thanks,  
Scott

---

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Sent:** Wednesday, May 24, 2023 9:44 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Subject:** RE: Question on new data sources requested for ISDW

Scott

Thank you. I am assuming this in the context of identifying the tool /technology then certainly We should ask. This would inform what kind of tool / capability is required for organization wide use. We can look for "Optimal" tool that is capable. Rolling out, integrating etc. can be opportunistic based on operating model. This way tool won't be limiting.

If it is for process, we can create a play book as we evolve the SDLC. Regardless, in my opinion the answer would inform us immensely.

Regards  
**Ananth**

---

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum

M 1-551-358-4609

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum  
M 1-551-358-4609  
[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

## Optum

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Wednesday, May 24, 2023 2:54 PM  
**To:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Subject:** RE: Question on new data sources requested for ISDW

I think the question below to ISDW could also be asked of EDP, SMART and DEEP. Do they routinely profile new candidate data sources before accepting them?

If yes, what do they use? If no, would they like to? Not sure if we want to consider that or just stick with DG land.

Scott

**From:** Barker, Scott D  
**Sent:** Wednesday, May 24, 2023 12:22 PM  
**To:** Ubele, Douglas J <[douglas\\_j\\_ubele@uhc.com](mailto:douglas_j_ubele@uhc.com)>; Kumar, Biren <[biren\\_kumar@uhc.com](mailto:biren_kumar@uhc.com)>  
**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Subject:** Question on new data sources requested for ISDW

Hi Doug and Biren,

Does ISDW have a standard practice to profile or otherwise QA new data sources that are being considered to be added to ISDW before they are added?

To make sure:

- data meets overall expectations
- critical fields have desired data quality
- get a handle on the amount of ETL scrubbing coding that may be needed

If you do some inspecting/profiling, what do you use?

Thanks,  
Scott

+++++

### RE: Lunch and Learn - Provider and Member MDM Overview

 Barker, Scott D  
To: ○ Ramakrishnan, Anantha; ● Cheema, Dave  
Cc: ○ Ferguson, Kimberly; ○ Paul, Sojan K  
Retention Policy: UHGlinbox (90 days) Expires: 8/22/2023  
Internal

Start your reply all with: [Thank you](#) [Yes, that makes sense.](#) [Good to know. Thanks.](#) [Feedback](#)

Just FYI.... There was a lunch and learn today for MDM. Every Sunday Member MDM runs compares between COMPAS and Member MDM to make sure Member MDM is accurate. So this would be another example of a monitoring (reactive) use case we have today.

**Member MDM** 

**What is Member MDM?**  
Member MDM is a Master Data Management tool that stores AARP members currently or previously insured in the SHIP program. Currently, this data is used for member validation during the claim intake process to determine matched and potential matched members and during the post adjudication.

**What data does Member MDM store?**  
Member MDM stores active and inactive member records containing Name, Addresses, Gender, Date of Birth, Date of Death, Medicare ID, Medicare Part A Effective Date, Medicare Part B Effective Date, AARP Membership Number, Day and Evening Phone Numbers, Email, Insured Plan Information, Auxiliary Person Information, COMPAS' Individual ID, Household ID, and Individual Termination Date.

**Where does Member MDM get data from?**  
Member MDM receives current and/or previously insured individuals from COMPAS. Currently, COMPAS is the source of truth for AARP Insurance Solutions members.

**Data Accuracy**  
Data accuracy is a component of data quality and is important for Member MDM to provide accurate member data to our clients. As Member MDM is not the source of truth now, we need to ensure the data we load in Member MDM matches the data received from COMPAS. The Data Accuracy job extracts data from COMPAS and Member MDM for all the attributes and completes a comparison on each individual attribute to identify any discrepancies. The job runs every Sunday at 10:00 PM EST to reduce the number of discrepancies due to timing.

Scott

**From:** Hoffman, Pamela <[pamela.hoffman@uhc.com](mailto:pamela.hoffman@uhc.com)>  
**Sent:** Wednesday, May 24, 2023 3:32 PM  
**To:** Kubler, Stephen R <[skubler@uhc.com](mailto:skubler@uhc.com)>; OCIO\_All\_IT Staff <[OCIO\\_All\\_ITStaff\\_DL@ds.uhc.com](mailto:OCIO_All_ITStaff_DL@ds.uhc.com)>  
**Cc:** Scott, Trevor <[trevor.scott@uhc.com](mailto:trevor.scott@uhc.com)>; Lisiecki, Lisa A <[lisa\\_lisiecki@uhc.com](mailto:lisa_lisiecki@uhc.com)>; Wilson, Cynthia <[cynthia.z.wilson@uhc.com](mailto:cynthia.z.wilson@uhc.com)>; Jackson, Jeffrey P <[jeffrey.p.jackson@uhc.com](mailto:jeffrey.p.jackson@uhc.com)>; Jones, Angela M <[a.jones@uhc.com](mailto:a.jones@uhc.com)>; Olsen, Joan M <[joan.m.olsen@uhc.com](mailto:joan.m.olsen@uhc.com)>; Tetzlaff, Michael W <[michael.w.tetzlaff@uhc.com](mailto:michael.w.tetzlaff@uhc.com)>; Lieb, Kathryn <[kathryn.lieb@uhc.com](mailto:kathryn.lieb@uhc.com)>; Herzog, Donna L <[donna.l.herzog@uhc.com](mailto:donna.l.herzog@uhc.com)>; Holenko, Rebecca L <[rebecca.holenko@uhc.com](mailto:rebecca.holenko@uhc.com)>; Fox\_Claim\_Product\_Owners <[BFF\\_Product\\_Owners\\_DL@ds.uhc.com](mailto:BFF_Product_Owners_DL@ds.uhc.com)>; Begum, Farheen <[fbegum@uhc.com](mailto:fbegum@uhc.com)>; Ziel, Jeff <[jeff.ziel@uhc.com](mailto:jeff.ziel@uhc.com)>; Packard, Jennifer <[jennifer.packard@uhc.com](mailto:jennifer.packard@uhc.com)>; Cook, BJ <[bj.cook@uhc.com](mailto:bj.cook@uhc.com)>; Collins, David <[david.collins@uhc.com](mailto:david.collins@uhc.com)>; Killian, Donna J <[donna.j.killian@uhc.com](mailto:donna.j.killian@uhc.com)>; Kniffin, Sabrina <[sabrina.kniffin@uhc.com](mailto:sabrina.kniffin@uhc.com)>; Norman, Twana A <[twana.a.norman@uhc.com](mailto:twana.a.norman@uhc.com)>; Ruth, Adrienne M <[adrienne.ruth@uhc.com](mailto:adrienne.ruth@uhc.com)>; McDowell, Scott E <[scott.e.mcowell@uhc.com](mailto:scott.e.mcowell@uhc.com)>; Ferguson, Courtney <[courtney.ferguson@uhc.com](mailto:courtney.ferguson@uhc.com)>; Poor, Amy L <[amy.poor@uhc.com](mailto:amy.poor@uhc.com)>; Sokoloff, Arthur <[arthur.sokoloff@uhc.com](mailto:arthur.sokoloff@uhc.com)>; Luckmann, Dawn M <[dawn.m.luckmann@uhc.com](mailto:dawn.m.luckmann@uhc.com)>; Hilbert-Baade, Genie M <[georgene.m.hilbertbaade@uhc.com](mailto:georgene.m.hilbertbaade@uhc.com)>; Vang, Sherry L <[sherry.vang@uhc.com](mailto:sherry.vang@uhc.com)>; IS\_TDM\_DataMasking <[IS\\_TDM\\_DataMasking@ds.uhc.com](mailto:IS_TDM_DataMasking@ds.uhc.com)>  
**Subject:** RE: Lunch and Learn - Provider and Member MDM Overview

Thank you everyone for attending today's Lunch and Learn. I hope you found it informative. If you have any questions regarding Provider and/or Member MDM, please reach out to our team.

**Subject:** RE: Lunch and Learn - Provider and Member MDM Overview

Thank you everyone for attending today's Lunch and Learn. I hope you found it informative. If you have any questions regarding Provider and/or Member MDM, please reach out to our team.

Thanks,  
Pam Hoffman

**Provider and Member MDM**  
Insurance Solutions - United Healthcare Medicare & Retirement  
680 Blair Mill Road, Horsham, PA 19044-2233  
E: [pamela\\_hoffman@uhc.com](mailto:pamela_hoffman@uhc.com) P: 215-902-8146

\*\*\*\*\*  
**RE: MRIS DQ project items that we would need to cover**

 Barker, Scott D  
To:  Cheema, Dave;  Ferguson, Kimberly  
Cc:  Ramakrishnan, Anantha;  Mandal, Aditya

Retention Policy: UHGLinbox (90 days)

Expires: 8/22/2023

 Internal

Start your reply all with: [Thanks for the feedback!](#) [This is great, thank you!](#) [See comments below.](#) [Feedback](#)

 Reply  Reply All  Forward

Wed 5/24/

Reviewing now so figured I'd write my thoughts down...

- For any tools that would be priced based on "seats", we would want to estimate 5 seats or so. We think tool would be mostly used just by us technicians
- I personally am skeptical about the Machine Learning in this context. A lot has to be put in place to make that fly.
- I would put dashboard functionality as low on list
- Highest priorities for me are 1, 3, 4 and 7
- 2 and 6 are wildcards for me – I can explain

Thanks,  
Scott

\*\*\*\*\*  
**RE: Queries regarding metadata lineage**

 Barker, Scott D  
To:  Mandal, Aditya  
Cc:  Ramakrishnan, Anantha;  Cheema, Dave

Retention Policy: UHGLinbox (90 days)

Expires: 8/21/2023

 Internal

 Reply  Reply All  Forward  

Tue 5/23/2023 12:50 PM

Hi, i think maybe this is what Omar can answer better when you meet with him and Raj and Srilatha. Omar can help us understand which reports use the columns in Raj and Sri's mapping sheet, and who uses those reports. That would be the final piece of the lineage.

But for the specifics below, I don't have much insight.

Thanks,  
Scott

**From:** Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>

**Sent:** Monday, May 22, 2023 1:58 PM

**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>

**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Subject:** Queries regarding metadata lineage

Hi Scott,

I have some queries regarding the use cases for which Metadata lineage are expected to be made.

Let's say for Smoking Surcharge- which Line Of Businesses are getting affected? (eg - Finance? Actuarial? Or any other??) Are there any reports regarding the Smoking Surcharge that comes out of Microstrategy that business users are interested? How do we get access to those reports?

Likewise similar info is needed for the other 2 use cases as well

1. A new plan is added to an existing member
2. Member paid a premium

Please guide us who would be the correct person(s) from whom can we get this info from? Having an understanding of this will help us create more value for the end users.

**Thanks & Regards,**

**Aditya Mandal**  
Sr. Consultant | Optum Advisory Services

M +91 981-864-0478  
[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)  
Noida, India

**Optum**

\*\*\*\*\*  
**Operationalizing Data Quality**

 Barker, Scott D  
Required:  Tannenbaum, Jim;  Hastie, Michael;  Searfass, David R;  Paul, Sojan K;  Ferguson, Kimberly;  Ramakrishnan, Anantha;  
 Cheema, Dave

Retention Policy: UHGLinbox (90 days)

Expires: 8/21/2023

 Accept  Tentative  Decline  Propose New Time  

Tue 5/23/2023 10:36 AM

 Please respond.

This meeting has been adjusted to reflect your current time zone. It was initially created in the following time zone: (UTC-05:00) Eastern Time (US & Canada). This appointment is next to another one on your calendar.

 Tuesday, May 30, 2023 1:00 PM-2:00 PM  Microsoft Teams Meeting

We would like to meet with the 3 of you to discuss the following topic.

This appointment is next to another one on your calendar.

Tuesday, May 30, 2023 1:00 PM-2:00 PM Microsoft Teams Meeting

We would like to meet with the 3 of you to discuss the following topic.

We are working with OPTUM Advisory Services to improve how we address data quality and data governance throughout the project lifecycle. We wanted to understand if/how/when the following areas are considered during the early stages of project intake. We met with Amanda Siebert from Portfolio Mgt, and she referred us to Jim/EA for aspects of project sizing early in the cycle. We are also meeting with David Weiss for once the project goes into Agile development.

During T-shirt sizing, are there user stories specific to Data Quality/Governance considered, for items such as:

- Will B2B's need to be written for new data sources?
- Will application utilize new data sources which should be profiled for expected data quality?
- Are data retention requirements being considered for the design?
- Will project use/store PII and PHI data, and therefore require masking of lower environments?
- Is time estimated for writing business and technical metadata?
- Will new application send out data extracts, and therefore require DUR (Data Usage and Rights) documents?

We look forward to meeting with you.

Thanks,  
Scott

\*\*\*\*\*  
Meeting 05/30/2023 with Scott and Kim F., Anantha R.

Subject: Global ID demo

- Background
    - Talked with Rose. The Group got disbanded. Then it stopped going forward. No PII, PHI. Subsequent upgrades took away some existing features.
  - What necessitated its replacement?
    - Never used it. Never tested it. Used internally. Nothing finalized.
    - This has not been used. Last time it ran was 2 years ago and the app was setup 4 years ago.
    - Not restarting it, the company is small; the product is unreliable; took a long time to implement any change. Even simple process took forever. Could not fix simple issues such as, sorting; made SMEs lose trust and confidence in it. It is buggy or would not even work.
    - Graphs are not very intuitive.
    - Has pretty dashboards that do not work.
    - Does not work well with Data Lineage and data retention.
    - Retention was given up due to issues.
    - Wanted it to be much more user friendly.
    - Not even mentioned in the Gartner's any magic quadrant.
    - Never even attempted to integrate it into the process.
    - No plans to renew the license.
  - In the Rules UI
    - Has SQL roles, execution history, you can write your own SQL
    - Currently it's not doing anything
    - It is just pulling metadata
  - Domain roles
    - This is where we see the roles
    - No implementation in streaming pipeline
- \*\*\*\*\*

## RE: DQ Operationalization Daily Standup

Barker, Scott D  
To: Ramakrishnan, Anantha; Cheema, Dave; Ferguson, Kimberly  
Cc: Mandal, Aditya; Paul, Sojan K  
Retention Policy: UHGlinbox (90 days)

Expires: 8/30/2023

Internal

Reply Reply All Forward ...

Thu 6/1/2023 11:25 AM

Hi,

I know Ananth has another meeting today at 1. Kim and Dave, let's take a checkpoint on where we are relative to schedule. I'd also like to increase Kim's involvement and reduce mine, where it makes sense.

ITEMS:

- Are we in week 3 or 4?
- Based on mtg yesterday with project mgt and today with EA, I think we are in good shape for SDLC change management aspects. We should:
  1. Get NFRs discussed today
  2. Set up working meetings with Stephanie and Dom from yesterday
- For DQ tool, I would like to have DQaaS demo, and talk to some current DQaaS customers
- For DQ tool, I think we need to be able to say we looked at some COTS products as due diligence.
  1. Confirm if Talend product (not OPEN Talend) is on PADU list
  2. Confirm if UHC/OPTUM has Talend enterprise license and how cost would compare to DQaaS
  3. speak with some UHC Talend customers if #2 is true

I think JET was correct in saying the home for all these data quality/data governance items is NFRs – Non Functional Requirements.

Thanks,  
Scott

## RE: DQ Operationalization Daily Standup

Barker, Scott D  
To: Ramakrishnan, Anantha; Cheema, Dave; Ferguson, Kimberly  
Cc: Mandal, Aditya; Paul, Sojan K  
Retention Policy: UHGlinbox (90 days)

Expires: 8/30/2023

Internal

Reply Reply All Forward ...

Thu 6/1/2023 1:07 PM

Notes:

- NFR's are the place, or one of the places, we want to document these DQ/DG items. It now seems so obvious!
- Scott tasks
  - gather more quality use cases to make sure we haven't missed any major "type". We have examples of simple and complex, proactive and reactive, but I want to talk to a few more people to see what we might have missed.
  - Find out what DEEP, SMART and EDP do today in terms of profiling new data sources.
- Kim/Dave tasks
  - Get current documentation on NFR's and review for DQ items
  - Set up one or two meetings next week with Stephanie and Dominic from yesterday to review how/when NFR's are evaluated and addressed in project planning and development (invite David W as optional, or copy on minutes)
  - Begin to document DQ items for NFR's – include whether item is mandatory or optional, and whether the project team does it, or engages EIM. For example, I would rather have EIM do profiling, but I think the metadata should be done by project with templates and assistance from DG. This gets us started on Roles and Responsibilities

- o Set up one or two meetings next week with Stephanie and Dominic from yesterday to review how/when NFR's are evaluated and addressed in project planning and development (invite David W as optional, or copy on minutes)
- o Begin to document DQ items for NFR's – include whether item is mandatory or optional, and whether the project team does it, or engages EIM. For example, I would rather have EIM do profiling, but I think the metadata should be done by project with templates and assistance from DG. This gets us started on Roles and Responsibilities
- o Move common objects to our SP site
- Ananth/Sojan tasks – we need to follow-up with Mike Hastie. Need some response on his mention about the open source tool he liked.

I thought meetings with David, Dave, JET were very positive.

Thanks,  
Scott

RE: DQ Operationalization Daily Standup

 Barker, Scott D  
 To: Cheema, Dave; Ramakrishnan, Anantha; Ferguson, Kimberly  
 Retention Policy: UHGlinbox (90 days) Expires: 8/31/2023  
 Internal  
 Start your reply all with: Sounds good, thanks! Sounds good. Sounds good to me. 

I have mtg at 1 which I forgot about. I think we are in good shape for week. Focus for next week can be writing up our checklist/assessment items, templates examples, etc. and mtg with Stephanie and others to see how we get these incorporated into NFR's – the perfect home. Also meeting with Mike Hastie.

On the tool side, would like to speak with a UHC DQaaS user, and a UHC Talend user.

I will work on the items 1.2, 1.3, 1.4 we discussed yesterday.

Scott

Review of progress with Liv (mostly Friday)

 Ramakrishnan, Anantha  
 To: Cheema, Dave; Mandal, Aditya  
 Cc: Malhotra, Dinesh  
 Retention Policy: UHGlinbox (90 days) Expires: 9/3/2023  
 Internal  
 Dave / Aditya

Had a 1:1 with Sojan, he liked what is done so far, thanks to all the hard work. Couple of things require critical thinking

- Sojan realizes the nebulous nature of Metadata lineage, he is expecting us to give a methodology and demonstrate it with couple of use cases. He is aware that it is not done before and business lineage is not thought out. He is expecting us to give them a guidance on it.
  - o **Follow up:** Come up with your template including each hop – See if you can bring out business aspect (in each hop) more than technical in each hop (We leverage all the technical work done by raj/Srilatha) – **Aditya : review preliminary work Tomorrow 06/06 in our touch point call**
- We will be reviewing the progress with Liv on Friday. So we need the following – **Dave - review of draft end of day 06/06**
  - o One summary slide – 4 streams of work (not at the task level, more summarized milestones) with two columns 1) expected outcome during analysis , 2) expected outcome post final delivery – ( This should directionally align with SOW / deliverables )
  - o One page for each of the effort – 4 quadrant slide covering ,
    - Summary of analysis and discovery,
    - Discussion and meetings held ( details with whom and bullet or two of synopsis)
    - Actions to be covered during the detailed execution phase
    - Any open item, follow ups, issues risks to be closed during execution
  - o Overview of work for next 3 week,
    - High level tasks – dates
    - next review – (proposed agenda)

Feel free to add anything you feel is missed.

Regards  
Ananth

RE: Kim, were you able to setup a meeting with Stephanie and Dom?

 Ferguson, Kimberly  
 To: Cheema, Dave  
 Cc: Barker, Scott D; Ramakrishnan, Anantha  
 Retention Policy: UHGlinbox (90 days) Expires: 9/3/2023  
 Internal  
 Start your reply all with: Thank you! Ok, thanks. Sounds good, thanks! 

Dom is tentative, but Stephanie is confirmed.

Thanks,

Kim

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Sent:** Monday, June 5, 2023 2:02 PM  
**To:** Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Cc:** Barker, Scott D <[scott.barker@uhc.com](mailto:scott.barker@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Subject:** Kim, were you able to setup a meeting with Stephanie and Dom?

Thank you.

Regards,  
**Dave Cheema**  
(952) 205-0802

RE: Intake flow document

 Tannenbaum, Jim  
 To: Cheema, Dave; Barker, Scott D; Paul, Sojan K; Ferguson, Kimberly; Ramakrishnan, Anantha; Mandal, Aditya  
 Cc: Searfass, David R  
 Retention Policy: UHGlinbox (90 days) Expires: 9/4/2023  




Tannenbaum, Jim

To: Cheema, Dave; Barker, Scott D; Paul, Sojan K; Ferguson, Kimberly; Ramakrishnan, Anantha; Mandal, Aditya

Cc: Searfass, David R

Retention Policy: UHGlinbox (90 days)

Reply | Reply All | Forward | | ...

Tue 6/6/2023 7:20 AM

Expires 9/4/2023

Internal

Start your reply all with: [Very helpful. Thank you!](#) [Thank you!](#) [Thanks. That helps.](#) [Feedback](#)

Hi All,

This is a typical intake document in Aha. [INSURANCE-I-552 Quick Pay - Alternative Email | Aha!](#) It gets discussed during a "KPI & Effort Assessment" meeting, where the creator describes the project in more detail (if necessary). At that point Harry Seiders tries to access cost-benefit and I give a high level (1 – 5 PIs) tech impact. The idea is then reviewed by a committee to determine priority. Once priority is assigned, we start the rest of the intake process.

Thanks,

Jet

---

## RE: Intake flow document



Ramakrishnan, Anantha

To: Barker, Scott D; Paul, Sojan K; Ferguson, Kimberly; Cheema, Dave; Mandal, Aditya

Cc: Shekhtman, Felix

Retention Policy: UHGlinbox (90 days)

Expires 9/4/2023

Reply | Reply All | Forward | | ...

Tue 6/6/2023 10:24 AM

Internal

Scott

Good morning. I think we have to look it from 2 angles.

- 1) If this at architecture level, I am wondering what NFR they consider and how does that get baked into ?
- 2) If at project level this becomes more functional for execution
- 3) Agree to all the details below.

Regards  
Ananth

---

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum

M 1-551-358-4609

[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

**Optum**

---

**From:** Barker, Scott D <[scott\\_barker @uhc.com](mailto:scott_barker @uhc.com)>  
**Sent:** Tuesday, June 6, 2023 8:32 AM  
**To:** Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>  
**Subject:** RE: Intake flow document

Hi,

I think that whatever we call these – NFRs or something else – we should create something along the lines of a "starter kit" that would contain:

- DG/DQ Checklist Items – B2Bs, DURs, profiling, metadata, etc.
- Detailed writeup of each item. answering the who/what/when/where/why questions for each
- Templates – like for metadata
- And more...

Let's discuss at our call today.

Thanks,  
Scott

---

**From:** Barker, Scott D  
**Sent:** Tuesday, June 6, 2023 9:18 AM  
**To:** Tannenbaum, Jim <[jet@uhc.com](mailto:jet@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>  
**Subject:** RE: Intake flow document

Jim/Dave,

Is there a place where NFR documents or templates are stored that we can view? Maybe we can use as a model or guide for our DG/DQ items.

Thanks,  
Scott

---

**From:** Tannenbaum, Jim <[jet@uhc.com](mailto:jet@uhc.com)>  
**Sent:** Tuesday, June 6, 2023 8:09 AM  
**To:** Barker, Scott D <[scott\\_barker @uhc.com](mailto:scott_barker @uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>  
**Subject:** RE: Intake flow document

Hey Scott,

I'm not sure we agreed. I believe that is correct. Most/all of the DG & DQ items feel like they are NFRs.

Thanks,

Jet

I'm not sure we agreed. I believe that is correct. Most/all of the DG & DQ items feel like they are NFRs.

Thanks,

Jet

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Monday, June 5, 2023 4:53 PM  
**To:** Tannenbaum, Jim <[jet@uhc.com](mailto:jet@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>  
**Subject:** RE: Intake flow document

Hi –

Just confirming... did we say that these aspects of data quality seem like non-functional requirements, and non-functional requirements are not looked at yet during intake. Is that right?

I agree with Jim's point that these DQ and DG items feel like NFR's, and so we would want to add these in the process wherever NFR's are assessed and planned for.

Thanks,  
Scott

---

**From:** Tannenbaum, Jim <[jet@uhc.com](mailto:jet@uhc.com)>  
**Sent:** Monday, June 5, 2023 1:03 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>  
**Cc:** Searfass, David R <[david\\_searfass@uhc.com](mailto:david_searfass@uhc.com)>  
**Subject:** FW: Intake flow document

Hi Team,

Please find the attached Process Intake Flow diagram. More in depth details are available, if needed.

Thanks,

Jet

---

**From:** McCarty, Michael J <[michael\\_mccarty@uhc.com](mailto:michael_mccarty@uhc.com)>  
**Sent:** Monday, June 5, 2023 12:47 PM  
**To:** Tannenbaum, Jim <[jet@uhc.com](mailto:jet@uhc.com)>  
**Subject:** RE: Intake flow document

Sure! See the attached document and let me know if you have questions. There's a more detailed version I can provide as well if this is too high level.

**Michael McCarty**  
Portfolio Planning Manager  
Insurance Solutions - United Healthcare Medicare & Retirement  
680 Blair Mill Road, Horsham, PA 19044-2233  
E: [michael\\_mccarty@uhc.com](mailto:michael_mccarty@uhc.com)  
P: 865-556-8901

---

**From:** Tannenbaum, Jim <[jet@uhc.com](mailto:jet@uhc.com)>  
**Sent:** Friday, June 2, 2023 8:27 AM  
**To:** McCarty, Michael J <[michael\\_mccarty@uhc.com](mailto:michael_mccarty@uhc.com)>  
**Subject:** Intake flow document

Hi Michael,

Would you please send me a copy of the project intake flow document? I'd like to share it with the data services team to illustrated how the pipeline currently works.

Thanks,

Jet

Jim Tannenbaum  
Sr Principal Engineer, TLCP  
UnitedHealthcare Medicare & Retirement Insurance Solutions  
680 Blair Mill Road  
Horsham, PA 19044-2233  
215-902-8553 (office) [jet@uhc.com](mailto:jet@uhc.com)  
**Our United Culture** The way forward  
■ Integrity ■ Compassion ■ Relationships ■ Innovation ■ Performance  
   

\*\*\*\*\*

Re: Talend Data Quality

 Townsend, Theresa M  
To:  Cheema, Dave  
Cc:  Ramakrishnan, Anantha;  Barker, Scott D;  Tyman, Mark A

Retention Policy: UHGlbox (90 days)

Expires: 9/5/2023

 Internal

Start your reply all with: [Thank you!](#) [Here is the application.](#) [Just submitted it!](#) [Feedback](#)

Wed 6/7/2023 12:56 PM

We requested the application via <https://dms.optum.com/>

---

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Date:** Wednesday, June 7, 2023 at 1:40 PM  
**To:** Townsend, Theresa M <[theresa\\_townsend@optum.com](mailto:theresa_townsend@optum.com)>  
**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Tyman, Mark A <[mark\\_tyman@optum.com](mailto:mark_tyman@optum.com)>

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Date:** Wednesday, June 7, 2023 at 1:40 PM  
**To:** Townsend, Theresa M <[theresa.townsend@optum.com](mailto:theresa.townsend@optum.com)>  
**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>, Barker, Scott D <[scott\\_barker\\_@uhc.com](mailto:scott_barker_@uhc.com)>, Tyman, Mark A <[mark\\_tyman@optum.com](mailto:mark_tyman@optum.com)>  
**Subject:** RE: Talend Data Quality

Theresa,

Thank you for accepting the meeting invite; looking forward to hear about the Talend from you.

I have a one more question: do you have any contact person at Talend that we could speak also? Any account rep. or technical contact will be a great help. We're also curious about the latest features of Talend Open Studio (Talend Data Quality). Thanks in advance.

Regards,

Dave Cheema

(952) 205-0802

---

#### RE: Data Governance New Project Toolkit



Ramakrishnan, Anantha

To:  Barker, Scott D;  Cheema, Dave;  Ferguson, Kimberly;  Mandal, Aditya  
Cc:  Paul, Sojan K

Retention Policy: UHGRinbox (90 days)

Expires: 9/5/2023

Internal

Reply Reply All Forward ...

Wed 6/7/2023 1:25 PM

Scott

Thank you. This can be very useful.

But do you think is the problem today , not being able to do these activities in structured way or not even thinking about them ?

I think we might want to tackle it in two folds

1. Primary Goal is to make DQ as practice , consider it as an essential "thing" for a completion and execution of project
2. Track that activity by KPI's to make sure projects has a place holder to evaluate DQ. the extent to which you have to incorporate DQ may vary. – Point being, DQ is necessary and the extent might vary from project to project
3. The aspects you had mentioned, - A play book of selected activities for doing it using a standardized practice and the guidelines.

Hope it makes sense.

Regards  
Ananth

---

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum  
M 1-551-358-4609  
[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

## Optum

---

**From:** Barker, Scott D <[scott\\_barker\\_@uhc.com](mailto:scott_barker_@uhc.com)>  
**Sent:** Wednesday, June 7, 2023 12:55 PM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>  
**Cc:** Paul, Sojan K <[sojan\\_paul@uhc.com](mailto:sojan_paul@uhc.com)>  
**Subject:** Data Governance New Project Toolkit

Hi,

Maybe we should have a Data Governance Project Lifecycle Toolkit – for those of us who may remember Ralph Kimball. ☺

In it we could describe all the activities, when they are done, who does, how to estimate for, etc. And provide links to templates and samples



### Sample Activities

Write B2B docs on new external data sources

- Describe activity, when it is done, who does, etc.
- Provide link to samples, template

Perform data profiling on new data sources

- Describe activity, when it is done, who does, etc.
- Provide link to samples, template

Write DUR docs on datasets that will be sent out

- Describe activity, when it is done, who does, etc.
- Provide link to samples, template

Understand and Design for Retention requirements

- Describe activity, when it is done, who does, etc.
- Provide link to samples, template

Follow DB naming standards

- Provide link to naming guidelines
- Provide link to standard naming abbreviations

Create Business and Technical Metadata

Follow DB naming standards

- Provide link to naming guidelines
- Provide link to standard naming abbreviations

Create Business and Technical Metadata

- Describe activity, when it is done, who does, etc.
- Provide link to samples, template

Thoughts?

Scott

+++++

**RE: data quality checks for new SMART sources**

 Barker, Scott D  
To:  Kakarlapudi, Verma V  
Cc:  Ferguson, Kimberly;  Shekhtman, Felix;  Cheema, Dave  
Retention Policy: UHGLinbox (90 days) Expires: 9/6/2023  


Thanks Verma.

Is it fair to say – some basic checks, but not to degree of items like below?

- Are phone numbers all 10 numbers, is there one single format, or many?
- Are special or hidden characters present in text fields?
- Are there critical amounts that are expected to be positive, but there are some 0 amounts, or negative amounts?
- A prospect received something before it was requested?
- Are there dates that don't make sense – like 01/01/1900
- Etc..

Thanks for your help Verma!

Scott

---

**From:** Kakarlapudi, Verma V <[verma\\_v\\_kakarlapudi@uhc.com](mailto:verma_v_kakarlapudi@uhc.com)>  
**Sent:** Thursday, June 8, 2023 2:34 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** Re: data quality checks for new SMART sources

Hi Scott,

In SMART we use ETL tool called DMExpress from Precisely. All inbound files go through data validation checks by this tool.

At file/record level

1. Directory/Filename for existence
2. Header/Trailer if applicable
3. Record Count (Detail records) against the count mentioned in Trailer
4. Validate number of fields in a record if file is not a Fixed width file

At field level

1. Datatype
2. Data Length
3. Not NULL if applicable
4. Value belongs to List of Values if applicable

A record gets written to an error file along with error message if any of the above conditions fail.

Please let me know if you have any further questions.

Thanks,

---

Verma V. Kakarlapudi

UnitedHealth Group | UnitedHealthcare Medicare & Retirement

680 Blair Mill Road | Horsham, PA 19044-2223

215.902.9061

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Thursday, June 8, 2023 10:49 AM  
**To:** Kakarlapudi, Verma V <[verma\\_v\\_kakarlapudi@uhc.com](mailto:verma_v_kakarlapudi@uhc.com)>  
**Cc:** Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** data quality checks for new SMART sources

Hi Verma,

Question for you...

As part of data governance and data quality, we are asking the following question:

- If your project or application will begin to consume a new data source (either from another group in UHG or from an external entity), what data profiling or data quality checking is performed to confirm the source data matches expectations in terms of content and quality?

For example, from speaking with Raj and Felix, we know that ISDW will do some ad hoc SQL checks on the important columns, but not a fuller "data profiling" exercise. If someone wants a new source added to SMART, what checks do you do to make sure the candidate source data meets data quality expectations? (I think we've all experienced getting new data and discovering surprises.)

If it's easier to have brief discussion, I can set up a call.

Hannu summer

I think we've all experienced getting new data and discovering surprises. ☺)

If it's easier to have brief discussion, I can set up a call.

Happy summer.

Thanks,  
Scott

++++++

### data quality checks for new EDP sources



Hi Sheharyar,

Question for you...

As part of data governance and data quality, we are asking the following question:

- If your project or application will begin to consume a [new data source](#) (either from another group in UHG or from an external entity), what data profiling or data quality checking is performed to confirm the source data matches expectations in terms of content and quality?

For example, from speaking with Raj and Felix, we know that ISDW will do some ad hoc SQL checks on the important columns, but not a fuller "data profiling" exercise. I'm not even sure how profiling would be done with less structured big data types. But if someone wants a new source added to EDP, what checks do you do to make sure the candidate source data meets data quality expectations? (I think we've all experienced getting new data and discovering surprises. ☺)

If it's easier to have brief discussion, I can set up a call.

Happy summer, and hope you're not impacted by our smokey weather conditions.

Thanks,  
Scott

++++++

### RE: OPTUM EDM PLAYBOOK 2022



Start your reply all with: [Unbelievable!](#) [Thanks for sharing!](#) [Very nice! Thank you!](#) [Feedback](#)

Read a few more emails for a surprise....

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Sent:** Thursday, June 8, 2023 9:39 AM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Cc:** Shekhtman, Felix <[felix.shekhtman@uhc.com](mailto:felix.shekhtman@uhc.com)>; Paul, Sojan K <[sojan.paul@uhc.com](mailto:sojan.paul@uhc.com)>  
**Subject:** RE: OPTUM EDM PLAYBOOK 2022

Thanks Scott., I remember seeing it sometime before 😊.

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum  
M 1-551-358-4609  
[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

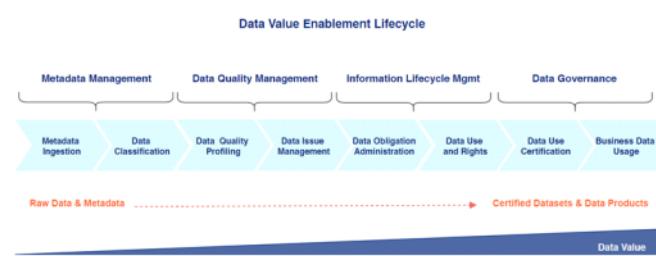
## Optum

**From:** Barker, Scott D <[scott.barker@uhc.com](mailto:scott.barker@uhc.com)>  
**Sent:** Wednesday, June 7, 2023 5:49 PM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Cc:** Shekhtman, Felix <[felix.shekhtman@uhc.com](mailto:felix.shekhtman@uhc.com)>; Paul, Sojan K <[sojan.paul@uhc.com](mailto:sojan.paul@uhc.com)>  
**Subject:** OPTUM EDM PLAYBOOK 2022

The old OPTUM EDM playbook done in 2020 was re-vamped in 2022. I think the new 2022 version is a very good document.

I was looking to see if it addressed how to [operationalize](#) stuff. It's more the theory and best practices, not so much how to operationalize.

But you'll smile a geeky smile when you read through it. We are very well aligned.





RE: List of data source owners

Barker, Scott D  
To: Ferguson, Kimberly; Cheema, Dave  
Cc: Ramakrishnan, Anantha

Retention Policy: UHGLinbox (90 days)

Internal

Some updates:

- I have meetings with Jenn (FOX product owner) and Sujaya (Batch Support) scheduled for next week for sample DQ use cases
- I sent notes to Verma/SMART and Sheharyar/EDP asking if they do any data profiling to determine quality for new data sources, what could be improved.
- We have 2 COMPAS use cases – Date of Death, and E-mail address

I should be able to round out with AoP and MDM next week.

A	B	C	D	E	F																				
Item #	Key Milestone	Start Date	End Date	Status	Comments																				
1.5	Identify data / Process owners	5/15/2023			<table border="1"> <thead> <tr> <th>Process</th> <th>Owner</th> </tr> </thead> <tbody> <tr> <td>MDM Quality Validation</td> <td>Pam Hoffman</td> </tr> <tr> <td>Fox Pre-batch quality checks</td> <td>Fox technical team</td> </tr> <tr> <td>SMART source data quality check</td> <td>Verma</td> </tr> <tr> <td>EDP</td> <td>Sheharyar</td> </tr> <tr> <td>DEEP</td> <td>Gurtej, Doug U.</td> </tr> <tr> <td>IDW</td> <td>Doug U.</td> </tr> <tr> <td>Fox QA</td> <td>Teena, Jenn</td> </tr> <tr> <td>COMPAS</td> <td>Adam Cimino</td> </tr> <tr> <td>AoP</td> <td>??</td> </tr> </tbody> </table> <p>Scott &amp; Kim will follow up</p>	Process	Owner	MDM Quality Validation	Pam Hoffman	Fox Pre-batch quality checks	Fox technical team	SMART source data quality check	Verma	EDP	Sheharyar	DEEP	Gurtej, Doug U.	IDW	Doug U.	Fox QA	Teena, Jenn	COMPAS	Adam Cimino	AoP	??
Process	Owner																								
MDM Quality Validation	Pam Hoffman																								
Fox Pre-batch quality checks	Fox technical team																								
SMART source data quality check	Verma																								
EDP	Sheharyar																								
DEEP	Gurtej, Doug U.																								
IDW	Doug U.																								
Fox QA	Teena, Jenn																								
COMPAS	Adam Cimino																								
AoP	??																								

Scott

Meeting 06/09/2023 with Dinesh Malhotra  
Subject: Review MRIS DQ progress and get review

My Notes:

- Ask Dinesh who to reach out to for Talend contact info.
- work with Sojan to reschedule the meeting
- Sojan does not like being late
- Have Scott lead it, let Scott do the talking for you
- Presentation should be very sharp
- No context slides
- Have some kind of time slider
- Focus on positive aspects
- We'll teach MRIS DQ how to do DQ
- Push meeting with Liv to Tuesday - done
- Scott has to be the spokesperson
- Consider losing the table, it loses whitespace, lose words;
- Lose discussions and meetings, instead use contributors - names & roles
- Analysis & discovery - what have we found, must be crisp of what we've discovered.
- Show Harvey balls in Progress of Outcome
- use icons of concerns
- Color code of what has been done
- Keep only 5 slides
- Anantha - prepare Scott

RE: Review / Guidance of Presentation

Barker, Scott D  
To: Ramakrishnan, Anantha; Paul, Sojan K; Cheema, Dave; Mandal, Aditya  
Cc: Malhotra, Dinesh

Retention Policy: UHGLinbox (90 days)

Internal

You replied to this message on 6/12/2023 1:54 PM.

Here are my intended talking points for tomorrow.

- I think overall we are at good place
- Enablement of DQ from Project Ideation/Intake → EA → Agile development
  - Folks we met with were all receptive – David W, Dave S, Mike H, Jim T, Amanda S
  - We have activities to include – mandatory profiling of new data, B2Bs, DUR's, metadata, retention, etc
  - Have to decide where some things go – is it NFR's (suggestion from Jim T), Solution Intent doc (suggestion from Dave S), beginning of project execution (Sprint 0?) - our goal is to shift proactive DQ approaches as far left as possible
  - We have yet to determine how we will track and make tollgates for projects – if we don't do this, our work will have "no teeth" and get put on a shelf
- Tool evaluation
  - Unless we discover some very unwelcome surprise, we are leaning to DQaaS – for obvious reasons.
  - We will have enough to demonstrate due diligence that we looked at Talend, INFA IDQ, and Open Metadata
  - We are working with OPTUM folks who manage GID servers. We will not renew GID contract, and we will re-use GID DB servers for "EIM"
  - We will salvage what we can from work over last 2 years with GID (AARP → SMART → DEEP) data lineage is an example. We will not attempt to "migrate" anything from GID to new tool.
- Formalize profiling of new data sources – either to OLTP systems, or analysis platforms. So far this appears more ad hoc than formalized.
  - New sources coming in for SDLC should be profiled
  - This would be prime use for DQaaS, in addition to ad hoc and "monitoring" type profiling

- Formalize profiling of new data sources – either to OLTP systems, or analysis platforms. So far this appears more ad hoc than formalized.
  - New sources coming in for SDLC should be profiled
  - This would be prime use for DQaaS, in addition to ad hoc and “monitoring” type profiling
- Understanding the overall footprint of High Impact Use Cases / Proactive vs. Reactive Example Use Cases
  - Member Over merge – working with Dave S and EIMP team we have:
    - Worked with COMPAS folks to reactively fix bad COMPAS data
    - Worked with COMPAS folks to determine additional business rule to be added to COMPAS to prevent incorrect Email scenario – still somewhat reactive
    - Supporting process analysis to see if there is truly “proactive” process change to eliminate problem altogether
  - Date of Death –
    - Example of DQ issue being remediated through automated process
  - “U Card” – this is a new emerging issue
    - U card is used by MA members, not MED SUPP
    - But we are seeing pattern where members who switch from Med Supp to MA plans are being impacted
    - We don’t yet know what, if any, our applications may be contributing to problem

Scott

## RE: Minutes of the meeting - DIG Assessment Team and Enterprise Analytics / reporting team meeting on 12.12.2022



Ferguson, Kimberly  
To: Barker, Scott D; Cheema, Dave  
Cc: Ramakrishnan, Anantha

Retention Policy: UHGlinbox (90 days)

Expires: 9/11/2023

Reply Reply All Forward

Tue 6/13/2023 9:35 AM

Internal

Start your reply all with:

Sure. I'll let you know what I come up with.

Thanks,

Kim

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Tuesday, June 13, 2023 9:55 AM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Subject:** RE: Minutes of the meeting - DIG Assessment Team and Enterprise Analytics / reporting team meeting on 12.12.2022

Hi Dave and Kim,

Last week I forwarded all the minutes from the sessions we had during the assessment phase. Please make sure you have read/reviewed these for representative DQ type issues and any other DQ insights or needs. These were the actual minutes – a bit more detailed than what went in final deck.

Also, I've attached a copy of DQ issues list from years past. Please review it as well for representative DQ issues and any other insights.

Kim – I have to focus on something else for next couple days. You have “carte blanche”. ☺

Thanks,  
Scott

**From:** Barker, Scott D  
**Sent:** Tuesday, June 6, 2023 10:06 AM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Subject:** FW: Minutes of the meeting - DIG Assessment Team and Enterprise Analytics / reporting team meeting on 12.12.2022

Please scan to see if includes specific data quality issues or scenarios.

**From:** Nand, Durga <[durga.nand@optum.com](mailto:durga.nand@optum.com)>  
**Sent:** Wednesday, December 14, 2022 5:03 PM  
**To:** Smith, Christopher B <[christopher.b.smith@uhc.com](mailto:christopher.b.smith@uhc.com)>  
**Cc:** Anantha Ramakrishnan, Srinivasan <[srinivasan.anantharamakrishnan@optum.com](mailto:srinivasan.anantharamakrishnan@optum.com)>; Malhotra, Dinesh <[dinesh.malhotra@optum.com](mailto:dinesh.malhotra@optum.com)>; Barker, Scott D <[scott.barker@uhc.com](mailto:scott.barker@uhc.com)>; Paul, Sojan K <[sojan.paul@uhc.com](mailto:sojan.paul@uhc.com)>  
**Subject:** RE: Minutes of the meeting - DIG Assessment Team and Enterprise Analytics / reporting team meeting on 12.12.2022

Hi Brian,

We would like to thank you for attending the meeting and providing your input in the assessment process. Attached are the minutes from that meeting. As we discussed in the meeting, Pl share the charter of data ideation team.

Please let me know if you have any question or concern and thanks once again for your help.

With Regards

Durga Nand | Optum  
Optum Advisory Services  
T +1 (952) 324-4007 (O)

\*\*\*\*\*

## RE: data quality checks for new EDP sources



Gulraiz, Sheharyar  
To: Barker, Scott D  
Cc: Shekhtman, Felix; Ferguson, Kimberly; Cheema, Dave

Retention Policy: UHGlinbox (90 days)

Expires: 9/11/2023

Reply Reply All Forward

Tue 6/13/2023 10:14 PM

Internal

Hi Scott,

It would be great to have a conversation on a call. Let's schedule a quick call to discuss.

Thanks

Thanks

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Thursday, June 8, 2023 10:44 AM  
**To:** Gulraiz, Sheharyar <[sheharyar\\_gulraiz@uhc.com](mailto:sheharyar_gulraiz@uhc.com)>  
**Cc:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** data quality checks for new EDP sources

Hi Sheharyar,

Question for you...

As part of data governance and data quality, we are asking the following question:

- If your project or application will begin to consume a new data source (either from another group in UHG or from an external entity), what data profiling or data quality checking is performed to confirm the source data matches expectations in terms of content and quality?

For example, from speaking with Raj and Felix, we know that ISDW will do some ad hoc SQL checks on the important columns, but not a fuller "data profiling" exercise. I'm not even sure how profiling would be done with less structured big data types. But if someone wants a new source added to EDP, what checks do you do to make sure the candidate source data meets data quality expectations? (I think we've all experienced getting new data and discovering surprises. ☺)

If it's easier to have brief discussion, I can set up a call.

Happy summer, and hope you're not impacted by our smokey weather conditions.

Thanks,  
Scott

FW: candidate data for DQ profiling

 Barker, Scott D  
To: Cheema, Dave; Ramakrishnan, Anantha  
Cc: Ferguson, Kimberly  
Retention Policy: UHGlinbox (90 days)  
Expires: 9/13/2023  
Internal  
You replied to this message on 6/15/2023 9:05 AM.  
Attachment: FOX\_PreBatch\_SQLs.txt (17 KB)

Reply | Reply All | Forward | ... | Thu 6/15/2023 7:11 AM

Dave,

Here are 10 DQ use cases where we are aware of data quality issues, but made decision to accept them and deal with consequences in work around.

All the queries in the attached are to find claims with data issues and withhold them from batch. They would cause batch to FAIL. The claims are then sent back to claim processor/whoever to be fixed, then get re-submitted to batch once fixed. So this can impact and delay claim processing and payment. These are all known bugs, we in the Agile methodology we chose to go ahead and release, and use work arounds until the root cause of the defect is fixed. There are 10 or so different issues being scanned for in the attached SQL.

So this can be written up as DQ issues that are known, and temporarily addressed with process work arounds (reactively), until the real fix can be done (proactively).

- Impacted Process – nightly batch
- Contact – Sujaya Linga/ Ops support
- Scenario – known DQ issues in new Fox Claims system. Decision made to take to production anyway, remediate reactively with people/process, until application code fix can be made to fix root cause
- Impact – delay to claim processing and payment. We are checking to see if delay is just one day, or can be more

Thanks,  
Scott

RE: DQ Operationalization Daily Standup

 Barker, Scott D  
To: Cheema, Dave; Ramakrishnan, Anantha; Ferguson, Kimberly  
Retention Policy: UHGlinbox (90 days)  
Expires: 9/21/2023  
Internal  
Start your reply all with: Sounds good, thank you | Sounds good, thank you | Thank you | Feedback

Hi Dave,

RE: DQ Operationalization Daily Standup

 Barker, Scott D  
To: Cheema, Dave; Ramakrishnan, Anantha; Ferguson, Kimberly  
Retention Policy: UHGlinbox (90 days)  
Expires: 9/21/2023  
Internal  
Start your reply all with: Sounds good, thank you | Sounds good, thank you | Thank you | Feedback

Thanks,  
Scott

Original Appointment --

From: Barker, Scott D  
Sent: Friday, May 19, 2023 3:08 PM  
To: Barker, Scott D; Cheema, Dave; Ramakrishnan, Anantha; Ferguson, Kimberly  
Cc: Mandal, Aditya; Paul, Sojan K  
Subject: DQ Operationalization Daily Standup  
When: Friday, June 23, 2023 1:00 PM - 1:30 PM (UTC-05:00) Eastern Time (US & Canada).  
Where: Microsoft Teams Meeting

Extending through June...

This is for daily stand-up call for next few weeks.

Included Aditya and Sojan as optional.

RE: DQ Operationalization Daily Standup

 Ramakrishnan, Anantha  
To: Barker, Scott D; Cheema, Dave; Ferguson, Kimberly  
Fri 6/30/2023 12:46 PM

## RE: DQ Operationalization Daily Standup

 Ramakrishnan, Anantha  
To: Barker, Scott D; Cheema, Dave; Ferguson, Kimberly  
Cc: Mandal, Aditya; Paul, Sojan K  
Retention Policy: UHGBInbox (90 days)

[Reply](#) [Reply All](#) [Forward](#)    
Fri 6/30/2023 12:46 PM  
Expires 9/28/2023

 Internal

May be 3 different pillars under Brian. Checks and balance.

Tracy - functional / business architecture – to ensure value is delivered  
Tom – Ensuring all the details are in place for execution / plan it  
David – execution

If it is the case, it is definitely Tom and Might be Tracy.

Regards  
**Ananth**

---

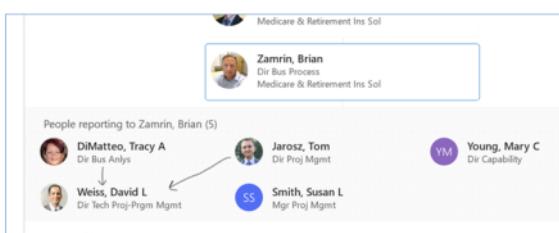
**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum  
M 1-551-358-4609  
[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

**Optum**

---

**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Friday, June 30, 2023 1:42 PM  
**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>  
**Cc:** Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>; Paul, Sojan K <[sojan.paul@uhc.com](mailto:sojan.paul@uhc.com)>  
**Subject:** RE: DQ Operationalization Daily Standup

Follow-up from call today.... Here is org chart snippet. Tracy has Business Architecture. Tom has project mgt – and if I heard correctly, it sounds like his group gets the project to somehow "complete requirements definition", then it goes to David's group purely for execution.



Scott

-----  
Sat 7/8/2023 9:49 AM

## RE: data quality checks for new EDP sources

 Barker, Scott D  
To: Gulraiz, Sheharyar  
Cc: Shekhtman, Felix; Ferguson, Kimberly; Ramakrishnan, Anantha; Cheema, Dave  
Retention Policy: UHGBInbox (90 days)  
Expires 10/6/2023  
 Internal

[Reply](#) [Reply All](#) [Forward](#)  

Thanks very much Sheharyar! This is very helpful for us to understand the overall pattern for consumption of big data sources, esp the raw to refined zone processing.

Ananth and Dave – I think our deliverable should have something about where/how we would envision perform "data quality" checks for unstructured data where at first there is no schema.

Kim – please put this in our project folder in SP.

Scott

---

**From:** Gulraiz, Sheharyar <[sheharyar\\_gulraiz@uhc.com](mailto:sheharyar_gulraiz@uhc.com)>  
**Sent:** Friday, July 7, 2023 9:45 PM  
**To:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Cc:** Shekhtman, Felix <[felix.shekhtman@uhc.com](mailto:felix.shekhtman@uhc.com)>; Ferguson, Kimberly <[kim.ferguson@uhc.com](mailto:kim.ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** RE: data quality checks for new EDP sources

Hello Scott,

I have created a detailed diagram that shows each zone in more detail. Please take a look at the attached document and let me know if you have additional questions.

I apologize for the late response.

Regards,

Sheharyar

**Sheharyar Gulraiz**  
Sr. Manager Data Engineering  
Insurance Solutions Enterprise Data Platform  
UnitedHealth Group | UnitedHealthcare Medicare & Retirement  
680 Blair Mill Road | Horsham PA, USA 19044-2233  
215.902.9036  
**Our United Culture. The way forward.**

Sr. Manager Data Engineering  
 Insurance Solutions Enterprise Data Platform  
 UnitedHealth Group | UnitedHealthcare Medicare & Retirement  
 680 Blair Mill Road | Horsham PA, USA 19044-2233  
 215.902.9036  
**Our United Culture. The way forward.**  
 ■ Integrity ■ Compassion ■ Relationships ■ Innovation ■ Performance

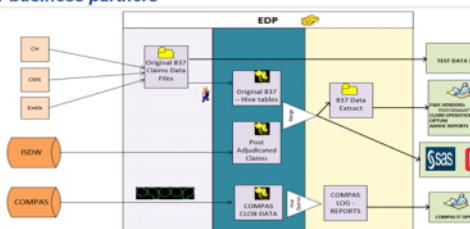
**From:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>  
**Sent:** Tuesday, June 27, 2023 9:26 AM  
**To:** Gulraiz, Sheharyar <[sheharyar\\_gulraiz@uhc.com](mailto:sheharyar_gulraiz@uhc.com)>  
**Cc:** Shekhtman, Felix <[felix\\_shekhtman@uhc.com](mailto:felix_shekhtman@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** RE: data quality checks for new EDP sources

Hi Sheharyar – thank you for your time the other week.

I pulled diagram below from the Data Services L&L we had a while back. If you have something that might elaborate on the "raw zone" and "refined zone" would you mind forwarding the doc or the link?

We're trying to understand conceptually where the activities would go for data quality checks and enriching the data (to fix any incoming data issues) in the big data type environment - where there is no "schema on write", and where HDFS files are write only. I'm old school relational world – so this is all new to me. ☺

### EDP is now comprised of several value offering solutions



Thanks,  
 Scott

FW: REVISED Weekly Portfolio Status Update - 07/07/2023

Barker, Scott D  
 To: Ferguson, Kimberly; Ramakrishnan, Anantha; Cheema, Dave  
 Retention Policy: UHGlinbox (90 days)  
 Expires: 10/6/2023 Sat 7/8/2023 11:21 AM  
 Internal

Here is this week's report from PPM group (Project Portfolio Mgt). Link at bottom takes you to pipeline.

Here are projects in the various stages from brand new to in execution. I'll try to look into further.

### Intake Pipeline

Last updated: less than a minute

Intake Pipeline						
New	In Review	Portfolio Priorities	Ready for Roadmapping	Ready for Execution (Planning)	Execution (Planning)	Execution (Delivery)
12   Innovative Plans   127 days	23   POS UW Incorporate UHC Med Supp Claims for Auto-adjustment   60 days	13   # USA MCO Claims Negotiations   16 days	# DTC OLE: Upload Documentation in OLE   240 days	# Oracle Exadata upgrade to v23   239 days	11   # RMO Process Risk Mitigation   57 days	1   # Shopping & Education - Rollout   236 days
19   New Legal Entity wave 5   127 days	22   SHIP Digital Onboarding Experience- Insured Member Website   16 days	15   # GI-OE: MS-MA-MS Trial and Work Eligibility   16 days	# Consolidated Signature/Disclosure Changes (Non-DTC OLE)   240 days	# RHEL Upgrade to v8 or higher UNICA-HCL Campaign   138 days	2   # DTC OLE 3.0 Redesign   157 days	3   # Customer Service Enhancements Program   240 days
A20 Program Support Needed   3 days	24   POS UW Incorporate Prior Applications for Auto-adjudication   15 days	21   # Real Time Submission API   52 days	# AAI AARP Membership API   16 days	# AAI AARP Membership API   16 days	4   # Multi Product Benefits   362 days	4   # Multi Product Benefits   362 days
Accept recurring credit card   131 days	26   Plan Change and Rate Quote Capability Expansion   30 days	25   # Member Portal Message Center/Secured Email   240 days	# Standardize Opt-in Language for all Channels   3 days	# Standardize Opt-in Language for all Channels   3 days	5   # CTPO Member SERVICING for Field Agents   240 days	6   # POS UW POS Manual UW for OLE acccs where IRIX
Accident Insurance in Agent Channel   127 days	APCD National Data Mart Feed   241 days	# DTIP: Agent Preference Center   206 days	# VA under 65			
Add Agent Name to COMPAS and xm file sent to Fidelity for Welcome Kit   18 days						
Add "Modified NHL" to NLE plans   127 days						

**From:** Savini, Lisa A <[lisa.savini@uhc.com](mailto:lisa.savini@uhc.com)>

**Sent:** Friday, July 7, 2023 3:31 PM

**Subject:** REVISED Weekly Portfolio Status Update - 07/07/2023

Revised Portfolio Highlights for week of 07/07/2023 (see revised projects Completion and Key Releases and Hot Topic updates).

UnitedHealthcare inspire. achieve. transform.

Insurance Solutions – Weekly Portfolio Status  
 As of July 7, 2023

Project Completion and Key Releases	
<ul style="list-style-type: none"> <li>RI - Enroll eligible individual under 65 - RI under 65 state requirements went live on 06/07/2023. However, on 06/09/2023, the state requested that we open an alternate enrollment period effective July 1, 2023, where all Medicare Supplement carriers must have a 60-day open enrollment period for all applicants under age 65 who are eligible for Medicare Part B due to disability or ESRD in which eligible individuals are able to enroll in Plan A. We are now live with this additional alternate enrollment period requirement as well.</li> <li>Commission Flexibility - PnP: ICM was released on 07/06/2023 and COMPAS &amp; ISDW were previously released on 05/05/2023 &amp; 05/23/2023 respectively. This initiative creates the ability to vary commissions by a percentage of the premium (PnP) for different age ranges.</li> </ul>	
Hot Topics	
<ul style="list-style-type: none"> <li>Projects at Risk – 2023 Top 10:           <ul style="list-style-type: none"> <li>Billing &amp; Collections Modernization (#3) – Milestone 2 work is 85% completed and Milestone 3 is 68% completed. The milestones and program are <b>Yellow</b>, at risk for scope/schedule/budget, and the team has completed building a Feature Burndown model that incorporates an evaluation on remaining Features required and their anticipated completion dates, as well as identified out-of-scope items. The team also completed</li> </ul> </li> </ul>	

<b>▲ Hot Topics</b>	
○ Projects at Risk – 2023 Top 10	
<ul style="list-style-type: none"> <li>○ Billing &amp; Collections Modernization (#3) – Milestone 2 work is 85% completed and Milestone 3 is 68% completed. The Milestones and program are <b>Yellow</b>, at risk for scope/schedule/budget, and the team has completed building a Feature Burndown model that incorporates an evaluation on remaining Features required and their projected completion dates, as well as identified get to green plans. The team also completed developing mitigation and contingency strategies which will entail confirmation/adjustment of plans and delivery dates and validation and the appropriate level of risk management planning. Focus continues to be put towards obtaining open required business decisions and approvals required for Feature development and the Program. The teams are closely monitoring and managing Sprint over Sprint Feature delivery and identified issues completing Features this prior Sprint that are actively being addressed. ISDW and MSTR reporting work streams are progressing according to plan. Data Migration detailed delivery plans are being further developed in the coming weeks and Parallel testing infrastructure setup progresses according to plan. An <b>end to end</b> Program Roadmap has been created that incorporates all activities related to Development, Testing, Training, Production Releases/Go Live, Operational Readiness and Operational Enablement.</li> <li>○ DTP: Member SERVICING for Field Agents (#5) - The project remains <b>Yellow</b> due to the amount of activity that could impact future releases. Project is off schedule due to development delays on the Jarvis team. The project health will be reevaluated closer to the July 19th release of 8 more tasks that will total 20 of 33 completed. <b>Exelsa</b> is continuing to deliver according to plan. A change request will be completed to account for the additional cost <b>Exelsa</b> has requested for the remainder of the project due to underestimation of the original project work. The project team is working towards a desired final delivery date by 09/01/2023.</li> </ul>	
○ Other Hot Topics:	
N/A	
<b>■ Project Approvals</b>	
<ul style="list-style-type: none"> <li>✓ Idea Expansion Approval           <ul style="list-style-type: none"> <li>○ N/A</li> </ul> </li> <li>✓ 2023 Priority Update:           <ul style="list-style-type: none"> <li>○ N/A</li> </ul> </li> </ul>	

#### Reporting Links:

(If desired Ahal access is restricted, please follow the detailed steps through [How to Request Ahal Viewer Access.docx](#))

Updated daily, Ahal dashboards are interactive user-friendly reporting tools which provide the functionality to analyze or interpret data in an illustrative experience.

Report Name <i>(To access report, please click on name)</i>	Report Description
<a href="#">Intake Pipeline</a>	The Intake Pipeline arranges projects based on the current ideation status. Status Update notes.

**Lisa Savini**  
Portfolio Management Office, Program Manager  
Insurance Solutions - United Healthcare Medicare & Retirement  
215-902-9060 | 215-902-8812 | [lisa.savini@uhc.com](mailto:lisa.savini@uhc.com)

680 Blair Mill Road | Horsham, PA 19044-2233

**OUR UNITED CULTURE** The way forward  
Integrity | Compassion | Relationships | Innovation | Performance



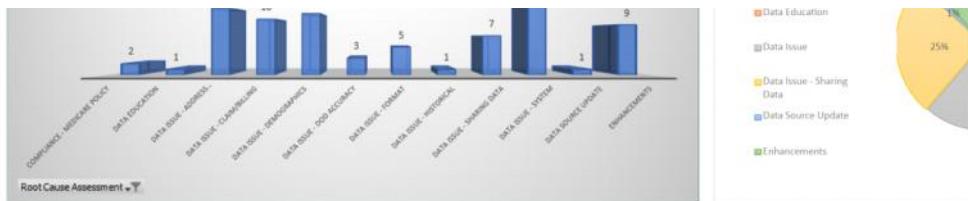
+++++  
Historical DQ report

Barker, Scott D  
 To ○ Ferguson, Kimberly; ○ Ramakrishnan, Anantha; ○ Cheema, Dave  
 Retention Policy UHGInbox (90 days) Expires 10/7/2023  
 Internal  
 Start your reply all with: [Thanks for the feedback!](#) [Thank you!](#) [Thanks, I will take a look.](#) [Feedback](#)

I went through historical DQ log. Some comments are:

- Saw issues we are dealing with today like DOD and MBI
- Saw range of issues from lower level, system specific to broader and across systems
- I think log and forum should be for problems that are higher impact, go across systems, have common themes, need escalation





## RE: REVISED Weekly Portfolio Status Update - 07/07/2023

Barker, Scott D  
To: Ferguson, Kimberly; Cheema, Dave; Ramakrishnan, Anantha  
Retention Policy: UHGLinbox (90 days)

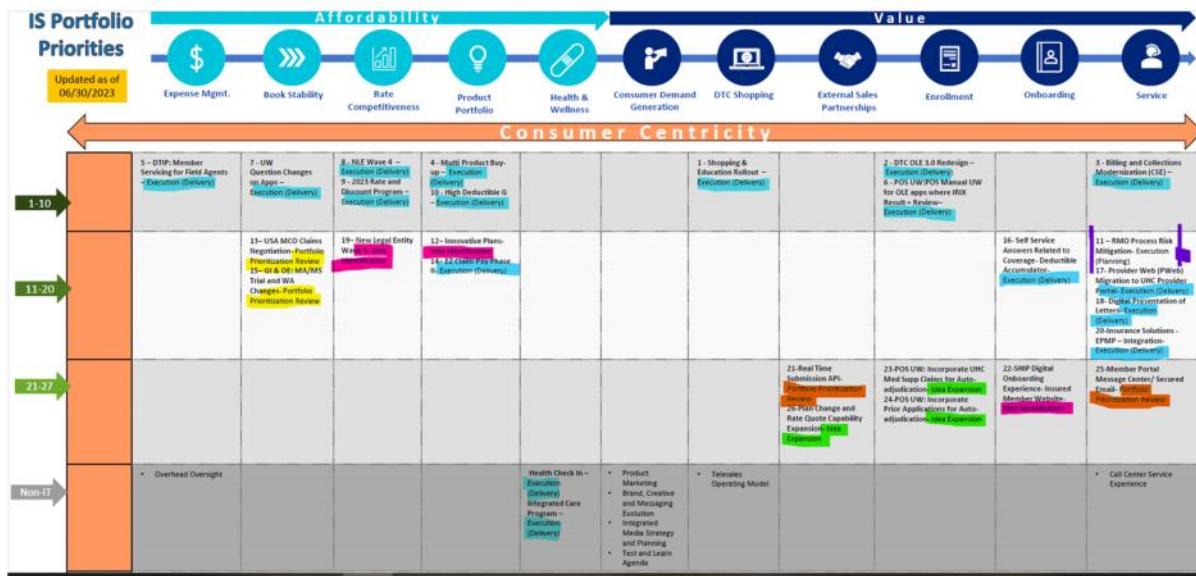
Expires: 10/7/2023

Internal

project intake and execution.png  
294 KB

Sun 7/9/2023 4:20 PM

It may not be pretty, but this shows where the projects are – most in delivery (blue), a few in portfolio prioritization review, few in idea expansion, one in execution planning, etc.



From: Savini, Lisa A <lisa.savini@uhc.com>

Sent: Friday, July 7, 2023 3:31 PM

Subject: REVISED Weekly Portfolio Status Update - 07/07/2023

Revised Portfolio Highlights for week of 07/07/2023 (see revised projects Completion and Key Releases and Hot Topic updates).

Insurance Solutions – Weekly Portfolio Status  
As of July 7, 2023

Project Completion and Key Releases	
<ul style="list-style-type: none"> <li>RI - Enroll eligible individual under 65 - RI under 65 state requirements went live on 06/07/2023. However, on 06/09/2023, the state requested that we open an alternate enrollment period effective July 1, 2023, where all Medicare Supplement carriers must have a 60-day open enrollment period for all applicants under age 65 who are eligible for Medicare Part B due to disability or ESRD in which eligible individuals are able to enroll in Plan A. We are now live with this additional alternate enrollment period requirement as well.</li> <li>Commission Flexibility - PdR: ICM was released on 07/06/2023 and COMPAS &amp; ISDW were previously released on 05/05/2023 &amp; 05/23/2023 respectively. This initiative creates the ability to vary commissions by a percentage of the premium (PdR) for different age ranges.</li> </ul>	
Hot Topics	
<ul style="list-style-type: none"> <li>Projects at Risk – 2023 Top 10: <ul style="list-style-type: none"> <li>Billing &amp; Collections Modernization (#3) – Milestone 2 work is 85% completed and Milestone 3 is 68% completed. The Milestones and program are yellow, at risk for scope/schedule/budget, and the team has completed building a Feature Burndown model that incorporates an evaluation on remaining Features required and their projected completion dates, as well as identified get to green plans. The team also completed developing mitigation and contingency strategies which will entail confirmation/adjustment of plans and delivery dates and validation and the appropriate level of risk management planning. Focus continues to be put towards obtaining open required business decisions and approvals required for Feature development and the Program. The teams are closely monitoring and managing Sprint over Sprint Feature delivery and identified issues completing Features this prior Sprint that are actively being addressed. ISDW and MSTR reporting work streams are progressing according to plan. Data Migration detailed delivery plans are being further developed in the coming weeks and Parallel testing infrastructure setup progresses according to plan. An end-to-end Program Roadmap has been created that incorporates all activities related to Development, Testing, Training, Production Releases/Go Live, Operational Readiness and Operational Enablement.</li> <li>DTIP: Member SERVICING for Field Agents (#5) - The project remains yellow due to the amount of activity that could impact future releases. Project is off schedule due to development delays on the Jarvis team. The project health will be reevaluated closer to the July 19th release of 3 more tasks that will total 20 of 33 completed. Exela is continuing to deliver according to plan. A change request will be completed to account for the additional cost. Exela has requested for the remainder of the project due to underestimation of the original project work. The project team is working towards a desired final delivery date by 09/01/2023.</li> <li>Other Hot Topics: N/A</li> </ul> </li> </ul>	
Project Approvals	
<ul style="list-style-type: none"> <li>Idea Expansion Approval <ul style="list-style-type: none"> <li>N/A</li> </ul> </li> <li>2023 Priority Update: <ul style="list-style-type: none"> <li>N/A</li> </ul> </li> </ul>	

### Reporting Links:

If desired, Atal access is restricted - please follow the detailed steps through [How to Request Atal Viewer Access here](#).

**Reporting Links:**(If desired Ahal access is restricted, please follow the detailed steps through [How to Request Ahal Viewer Access.docx](#))

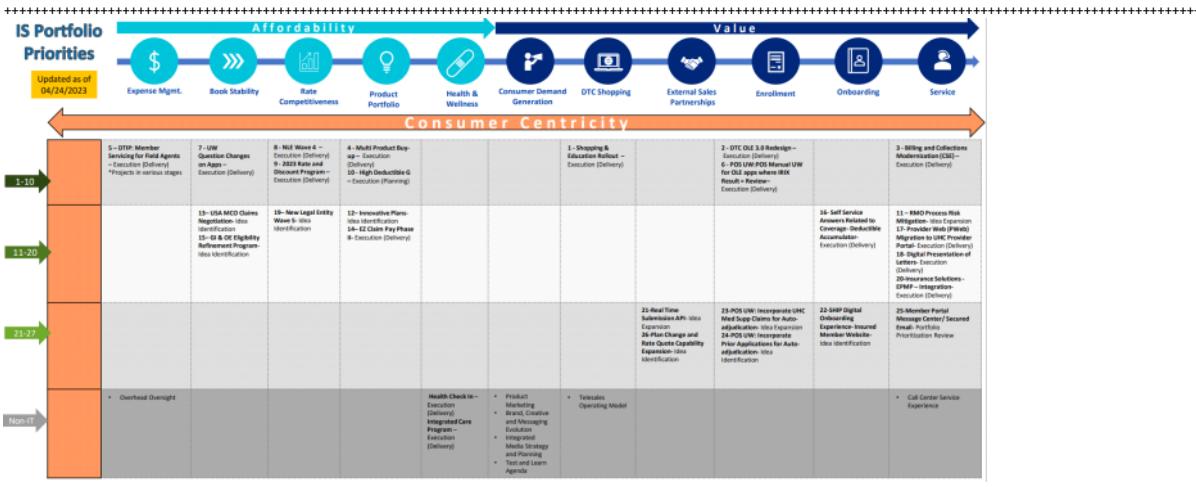
Updated daily, Ahal dashboards are interactive user-friendly reporting tools which provide the functionality to analyze or interpret data in an illustrative experience.

Report Name (To access report, please click on name)	Report Description
<a href="#">Intake Pipeline</a>	The Intake Pipeline arranges projects based on the current ideation status, number of days spent in that status, and Priority number.
<a href="#">Go Live Report</a>	The 'Go Live' dashboard arranges projects, based on planned go-live date and displays risk status, in a monthly timeline. Users can identify the project name, view its planned go-live date/Health status, and narrow in on a project's Strategic Choice, Go Live Date, and most recent Status Update notes.

**Lisa Savini**  
**Portfolio Management Office, Program Manager**  
 Insurance Solutions - United Healthcare Medicare & Retirement  
 ☎ 215-902-9066 ☐ 215-902-8812 ☐ [lisa.savini@uhc.com](mailto:lisa.savini@uhc.com)

680 Blair Mill Road | Horsham, PA 19044-2233

**OUR UNITED CULTURE** The way forward  
 Integrity | Compassion | Relationships | Innovation | Performance



RE: Informatica | Optum Insights - Data Quality Demo

Scholle, David <dscholle@informatica.com>  
 To ☎ Ramakrishnan, Anantha; ☎ Cheema, Dave  
 Cc ☎ Watson, Richard

Retention Policy UHGLinbox (90 days)

Expires 10/10/2023

Reply Reply All Forward

Wed 7/12/2023 5:07 PM

Internal

Suggested Meetings

Get more add-ins

Anantha and Dave,

Below is directional IDMC pricing based on the information provided, "Optum Discount" column. I'd like to mention our Informatica processing unit (IPU) pricing scales down as the annual IPU usage increases.

There are significant price break points at 300, 500, 1000, 2000, 3000, 5000 and so on. 80 is basically the entry point of IPU and as you add IPU's price will decrease. Also, you can share IPU's within an ORG by adding sub orgs. I'd like to explain the business value of this element of our IPU pricing further.

Also I'd like to get more detail on your exact use case and be sure we are directionally accurate— and would like to ask for 15 meeting to discuss. Please let me know a few times that work for you.

There are significant price breaks for you, and, today, there's more, more and so on. So I'm looking into some pricing for you and it's a price to decrease. Also, you can share IPU's within an ORG by adding sub orgs. I'd like to explain the business value of this element of our IPU pricing further.

Also I'd like to get more detail on your exact use case and be sure we are directionally accurate— and would like to ask for 15 meeting to discuss. Please let me know a few times that work for you.



**From:** Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>  
**Sent:** Monday, July 10, 2023 3:42 PM  
**To:** Scholle, David <[dscholle@informatica.com](mailto:dscholle@informatica.com)>  
**Subject:** FW: Informatica | Optum Insights - Data Quality Demo

**Richard Watson | Informatica**

**Healthcare Information Governance**

**End-to-End Data Management**

Informatica [privacy policy](#) | Email [rwatson@informatica.com](mailto:rwatson@informatica.com)

[unsubscribe](#)

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Sent:** Monday, July 10, 2023 3:41 PM  
**To:** Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>  
**Cc:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** RE: Informatica | Optum Insights - Data Quality Demo

Hi Richard

Greetings. We need help from you. We need some directional pricing information to make next level of decision with top management. We need pricing for 2 options

- 1) On Prem – IDQ – 2 Crores / 15 users / might be 500GB of data gets analyzed a month
- 2) Cloud option

Please let us know what is the support level and if there are additional cost for enhanced support. Again we just need indicative directional pricing for planning / budgeting needs. Not accurate, that we can discuss later.

We would like to get this by tomorrow so it can be baked into our next round review.

Regards  
Ananth  
-----

**From:** Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>  
**Sent:** Wednesday, June 28, 2023 3:35 PM  
**To:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Cc:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** RE: Informatica | Optum Insights - Data Quality Demo

Hi Ananth,

No problem – can we do 10am CT tomorrow for 30 minutes?

**Richard Watson | Informatica**

**Healthcare Information Governance**

**End-to-End Data Management**

Informatica [privacy policy](#) | Email [rwatson@informatica.com](mailto:rwatson@informatica.com)

[unsubscribe](#)

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Sent:** Wednesday, June 28, 2023 12:27 PM  
**To:** Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>  
**Cc:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Subject:** RE: Informatica | Optum Insights - Data Quality Demo

Hi Richard

Apologies I missed your call yesterday. I can meet tomorrow afternoon or Friday . If that is not feasible we have to push to 07/06, which is not ideal.

Regards  
Ananth

---

**Anantha Ramakrishnan**  
Sr. Director, Digital Transformations | Optum  
M 1-551-358-4609  
[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

**Optum**

**From:** Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>  
**Sent:** Tuesday, June 27, 2023 11:58 AM  
**To:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>  
**Subject:** RE: Informatica | Optum Insights - Data Quality Demo

Hi Ananth,

Just bumping this up for visibility. Could we connect tomorrow or next week on the 6<sup>th</sup> or 7<sup>th</sup>?

**Richard Watson | Informatica**

Hi Ananth,

Just bumping this up for visibility. Could we connect tomorrow or next week on the 6<sup>th</sup> or 7<sup>th</sup>?

**Richard Watson | Informatica**

**Healthcare Information Governance**

**End-to-End Data Management**

Informatica [privacy policy](#) | Email [rwatson@informatica.com](mailto:rwatson@informatica.com)

[unsubscribe](#)

**From:** Watson, Richard

**Sent:** Monday, June 26, 2023 11:26 AM

**To:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>; Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Cc:** Scholle, David <[dscholle@informatica.com](mailto:dscholle@informatica.com)>

**Subject:** RE: Informatica | Optum Insights - Data Quality Demo

Hi Ananth,

Yes that would be great. Could we connect for 30 minutes this afternoon or later this week to set an agenda as you suggested?

Thanks,

**Richard Watson | Informatica**

**Healthcare Information Governance**

**End-to-End Data Management**

Informatica [privacy policy](#) | Email [rwatson@informatica.com](mailto:rwatson@informatica.com)

[unsubscribe](#)

**From:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>

**Sent:** Monday, June 19, 2023 2:25 PM

**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>; Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>

**Subject:** RE: Informatica | Optum Insights - Data Quality Demo

Richard

If you want to have quick connect to brainstorm the agenda, points to cover etc., we are open and can spend 30 mins to prepare. Please let us know.

Regards  
**Ananth**

---

**Anantha Ramakrishnan**

Sr. Director, Digital Transformations | Optum

M 1-551-358-4609

[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)

**Optum**

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Sent:** Monday, June 19, 2023 1:42 PM

**To:** Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>

**Cc:** Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>

**Subject:** RE: Informatica | Optum Insights - Data Quality Demo

Richard,

Thank you for putting a demo session together. However, the leadership felt that the team was not well informed about our platform, our problem domain, and our specific needs, instead they were focused on the cloud and that is not specific to our problem domain.

So here is the context: our platform is strictly on-prem and we have Data Quality issues. We had understood that there is an Informatica Data Quality tool, which peaked our interest. We would like to request the following: could you please put together a team of experts who know how best to do the data quality for the on-prem (please no cloud sellers). And how to put an optimum solution together, meaning which products, connectors, and architecture will best meet our needs? Even though our long-term plan is to move to the cloud, but not in the immediate future. If you can arrange that meeting sometime this week, that'll be great.

I hope that clarifies our expectations. But if you still have any questions/concerns, please do not hesitate to contact me. Thank you.

Regards,

**Dave Cheema**

(952) 205-0802

-----Original Appointment-----

**From:** Cheema, Dave **On Behalf Of** Watson, Richard

**Sent:** Wednesday, June 14, 2023 11:38 AM

**To:** Ramakrishnan, Anantha; Mandal, Aditya; Malhotra, Dinesh; Shin, John S; Koneti, Rajesh; Nand, Durga; Tee, Vicky Hui Meng; Barker, Scott D; Ferguson, Kimberly; Paul, Sojana K; Palla, Sandeep

**Subject:** FW: Informatica | Optum Insights - Data Quality Demo

**When:** Thursday, June 15, 2023 3:00 PM-4:00 PM (UTC-06:00) Central Time (US & Canada).

**Where:** Microsoft Teams Meeting

All,

This demo is scheduled for Informatica to showcase its Informatica Data Quality tool, its features and answer any questions you may have. In addition, we would like you to fill out the template listed below and provide **User Rating** for each feature from your perspective. If you would like to provide any additional information, please use the **Comments** column.

Item #	Category	Feature	Description	User Rating (1-5)	Comments
1	Tool functionality	Data Profiling and Assessment	The ability of a DQ to collect and analyze data to understand its quality, structure and		

Item #	Category	Feature	Description	User Rating (1-5)	Comments
1	Tool functionality	Data Profiling and Assessment	The ability of a DQ to collect and analyze data to understand its quality, structure and content. It can help identify data issues, such as, missing values, duplicate records, out-of-range values, and inconsistent formats		
2	Tool functionality	Data Cleansing and Standardization	Identify incomplete/incorrect/irrelevant data. Then replace/modify/delete dirty data		
3	Tool Functionality	Data Types	Can the DQ tool to handle data types such as, structured, semi-structured and unstructured data		
4	Tool functionality	Data Sources	The tool should be able to handle data sources such as, databases, files, and web services		
5	Tool functionality	Data Quality Rules and Monitoring	The ability to define DQ rules for acceptable data quality		
6	Advanced Functionality	Machine Learning and Automation	ML can be leveraged for such capabilities as Anomaly Detection, Prediction of Missing Data, Continuous Learning and Adaptability, etc.		
7	Integration	Data Integration and Connectivity	Ability to integrate with other data sources and systems, such as, well-known databases, data warehouses, and data lakes.		
8	Integration	Data Governance and Metadata Management	The tool should support data governance practices; provide features to define data quality rules, implement data policies, and regulatory requirements.		
9	Adaptability and self-serve	Ease of Use and User Interface	The tool's user interface and overall user experience should be intuitive, easy to navigate, and provide a user-friendly environment for technical and non-technical users		
10	Adaptability and	Data Quality	Visual representation of data quality		

#### RE: Informatica | Optum Insights - Data Quality Demo

 Scholle, David <dscholle@informatica.com>  
 To: ○ Ramakrishnan, Anantha; ● Cheema, Dave  
 Cc: ○ Watson, Richard  
 Retention Policy: UHGlinbox (90 days)

[Reply](#) [Reply All](#) [Forward](#)    
 Wed 7/12/2023 5:07 PM



Suggested Meetings

+ Get more add-ins

12	Affordability	Licensing on-prem vs. cloud	Does the tool require different licenses for on-prem and cloud? Could have financial impact		
13	Affordability	Transferability of license	Can the license be transferred from one organization to another? This could be due to reorg/acquisition/merger, etc. Could have financial impact and service disruption		

Please feel free to forward this invite to anyone you deem appropriate. Do not hesitate to contact me, should you have a question/concern. Thank you.

Regards,  
Dave Cheema

-----Original Appointment-----

**From:** Watson, Richard <[rwatson@informatica.com](mailto:rwatson@informatica.com)>  
**Sent:** Monday, June 12, 2023 1:35 PM  
**To:** Watson, Richard; Sotebeer, Greg; Scholle, David; Cheema, Dave  
**Cc:** Moser, Christopher; Banik, Sukanya  
**Subject:** Informatica | Optum Insights - Data Quality Demo  
**When:** Thursday, June 15, 2023 3:00 PM-4:00 PM (UTC-06:00) Central Time (US & Canada).  
**Where:** Microsoft Teams Meeting

Please confirm receipt.

Meeting ID: 290 552 050 341

Passcode: V4Rebz

[Download Teams](#) | [Join on the web](#)

[Learn More](#) | [Meeting options](#)

This e-mail, including attachments, may include confidential and/or proprietary information, and may be used only by the person or entity to which it is addressed. If the reader of this e-mail is not the intended recipient or intended recipient's authorized agent, the reader is hereby notified that any dissemination, distribution or copying of this e-mail is prohibited. If you have received this e-mail in error, please notify the sender by replying to this message and delete this e-mail immediately.

\*\*\*\*\*  
**RE: Who adds items to Technical Debt or IRAADs**

 Byrd-Harrell, Stephanie B  
 To: ● Cheema, Dave; ○ Fong, Dominic H  
 Cc: ○ Barker, Scott D; ○ Ferguson, Kimberly; ○ Ramakrishnan, Anantha  
 Retention Policy: UHGlinbox (90 days)

[Reply](#) [Reply All](#) [Forward](#)    
 Wed 7/26/2023 8:13 PM



Start your reply all with: [Thank you for the clarification.](#) [It does. Thank you!](#) [Very helpful. Thank you!](#) [Feedback](#)

Hi Dave



Start your reply all with:

[Thank you for the clarification.](#)

[It does. Thank you!](#)

[Very helpful. Thank you!](#)

[Feedback](#)

Hi Dave.

Definitely our Project Managers for the Project Issues and Risks and Decisions.  
We maintain a global IRAAD which the Proj Managers enter into for their assignments.

Tech Debt is the responsibility of the System Architects. They manage that outside of the PI Planning often.  
Remember that the Project Managers are "beholden to" the Business Owner, so they are capturing issues, risks and decisions which block on time/on scope delivery.

Hope this helps!

Steph

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Sent:** Wednesday, July 26, 2023 8:18 PM

**To:** Byrd-Harrell, Stephanie B <[stephanie\\_byrd-harrell@uhc.com](mailto:stephanie_byrd-harrell@uhc.com)>; Fong, Dominic H <[dominic\\_fong@uhc.com](mailto:dominic_fong@uhc.com)>

**Cc:** Barker, Scott D <[scott\\_barker@uhc.com](mailto:scott_barker@uhc.com)>; Ferguson, Kimberly <[kim\\_ferguson@uhc.com](mailto:kim_ferguson@uhc.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>

**Subject:** Who adds items to Technical Debt or IRAADs

Steph, Dominic,

Do you know who adds items to the Technical Debt or IRAADs? Is it your group or somebody else? If somebody else, would you happen to know who might that be? Thanks in advance.

# Meetings

Wednesday, April 5, 2023 2:24 PM

Meeting 04/05/2023 with DQaaS

Subject:DQaaS

Internal capability

Run locally in Kubernetes environment, can it run on a stand-alone server

Database access can be an issue, when you don't have proper access

Who is responsible for the operations and maintenance

what is costing model - added to a group, per tenant - \$8771/month/tenant

\*\*\*\*\*  
Meeting 05/17/2023 with Anantha

Subject: Daily catch-up

what are we doing today for DQ?

Give the situations and frequencies that come up

Create DOCUMENT to analysis and discoveries

How often are they being done?

Which team is facing these issues

Jim Tannenbaum (EA), Mike Hastie (EA)

David Weiss (SDLC, PMO)

Agent database for use cases

\*\*\*\*\*  
Meeting 05/24/2023 with Anantha

Subject: DQ Features current state

Operation of Data Quality in the cloud

How does the licensing work on-prem vs. cloud?

Data quality built into operational model

Want some sort of POC (very high level)

\*\*\*\*\*  
**Meeting 6/1/2023 - Subject: SDLC Architecture Meeting**

Operationalization of Data Quality

Implement DQ from Ideation to operations

Sojan - to reduce member impact and raise confidence in data, data retention, and metadata

New data sources should have B2B document, they should be profiled. Data retention should be part of the DQ

Respect PII and PHI

Implement data governance for outgoing data

Our portfolio intake process starts from Ideation

Bring an analyst and architect to expand ideation

Develop high level plans, then goes to CBA and present high level estimates

After approval, get the next level of design; eventually project goes into details

They T-shirt size at this level only

Most of Scott's ideas and fall into NFRs, they are part of the execution

Should the DQ be handled by DQ team, a better way will be to embed it into the dev team

It is not part of the process today

DQ impact due to the lack of DQ, is the impact factored? If yes, how?

They use templates and review and revise EA templates

For example, consumer experience must not be degraded

There are no EA tollgates at present

We will have to figure out which checks belong where

We need to elevate DQ leveraging architecture

\*\*\*\*\*  
Meeting 06/07/2023 with Dinesh

Subject: client interaction guidance

Liv wants status bi-weekly

one tab for each tool

create an executive summary

setup meetings with DQ vendors and invite Dinesh's team

work with Sojan

Sojan does not like being late

Have Scott lead it, Scott doing the talking for you

Presentation shuld be very sharp

No context slides

Have some kind of time slider

Focus on positive aspects

We'll teach MRIS DQ how to fish

Push meeting with Liv to Tuesday

Scott has to be the spokesperson

Dinesh: consider losing the table, it loses whitespace, lose words;

Lose discussions and meetings, instead use contributors - names & roles

Analysis & discovery - what have we found, be crsip of what we've discovered. first b

Show Harvey balls in Progress of Outcome

use icons of concerns

Color code of what has been done

Keep only 5 slides

Anantha - prepare Scott

\*\*\*\*\*  
Meeting 06/09/2023 with

Meeting 06/09/202 with Jim Tannenbaum (JeT) and David Searfass

Subject: MRIS DQ Process Intake Process

It starts with Process Intake template being submitted in Aha!

The only field required is the name of the idea

Then Intake --> prioritization --> Execution

Kieffer would be involved at the intake

The idea gets assigned to an architect to produce intake process

From Execution, estimates come during execution phase

Architect takes it to the DUR team

Need approval to access the document

Most of it happens at execution

Architecture is at the end of the phase

For reports, talk to Omar

EMR --> Sojan

Tech team decide the data flow

M360 --> B360 is the Optum platform

It may deviate the flow based on technology, e.g., DEEP, DW etc.

99% of the time, idea goes to solution

Regulatory requirement may have special requirements

EA/BA do the solution

After solution, do the solution doc

COMPAS team is good at B2B

Execution is responsible for the NFRs

ISB and EA are subject to more rigor

Planning Interval (PIs) happen at the system level

Architecture Solution Intent document would be the ideal place to add DQ as a success criteria

The purpose of this template/document is considerations and influence solution

IDX is the business

HPC engineering practices

Architecture considerations

\*\*\*\*\*

Meeting 06/15/2023 with Dinesh

Subject: Status update

Collected use cases, learned where it would make sense to inject them

DQ tool demos are scheduled

SDLC process DQ entry points and defining success criteria are identified

Now we've starting templates, guidelines

Reschedule this meeting to accommodate Aditya

Dinesh want to see the score card

Define RACI to all functions - make ownership clear

Architecture team should define the success criteria

Give them mock template

Lineage origination could be someplace else

\*\*\*\*\*

Meeting 06/15/2023 with Talend

Subject: Talend Data Quality vendor

Only in commercial version

In it a service, or an executable that run client environment

Different licenses for on-prem and cloud

Licenses are based on number of users

\*\*\*\*\*

Meeting 06/20/2023 with Talend

Subject: Demo

Meeting 06/20/2023 Talend

Subject: Demo

**Product**

Talend Data platform is complete, modular, flexible and trusted data platform

Can run on on-prem, cloud, multi-cloud and hybrid platforms

Inside Talend Data Fabric, it has capabilities such as, Data Catalog, Data Inventory

Data Inventory - web based capability, creates data quality metrics, it is the launching point

Data quality is Talend's core capability

Data Integration is data quality in the MDM manner

Data Stewardship is to resolve complex rules and data cleansing issues

**Demo**

Login and see the dashboard with things you're allowed and show DQ dimensions

Dashboard rolls up all dimensions with some more details and suggestions, where needed. It does triage of data quality

You can view the DQ rules in the dashboard as well

Discoverability shows data tags

A dataset is a profile of data with a bunch of metadata attached to it. Dataset is a file in the file system

Profiling view is context sensitive view, where you can see multiple levels of details

Most of error detections and profiling is out of the box. It make remediation recommendations

We can sample initial rows or random rows. MongoDB data stores metadata

Architecture

Central data management is directing and other groups are using it

It can integrate with other tools and datasets

Any third party that has API interface, Talend can work with it

Remote engine runs in your VPC and the data does not move  
No throttling in the remote engines  
You can move your job to other available environments without any additional cost  
Remote engine can be sold as a separate job in the marketplace

#### **Connectivity**

Can the DQ be added to the ETL offering  
DQ job can be called from external API

#### **Training**

Several levels of support, standard business, standard business week, mission critical, etc. available

#### **Education**

Different levels of support, to setup use their professional services. The support services enable you to be self-sufficient, and also have online academy.

Good presentation - mixture of technology and business

#### **UHG**

UHG is a part of Talend's Strategic sales model  
UHG is part of the commercial and open source license  
No standard service agreement available today  
UHG does not want to own the infrastructure

+++++

Meeting 06/30/2023 with EA (David Searfass)

Subject: review of DQ enablement in EA/SDLC

David S. trying to understand whether infuse into current templates or create a separate document

Business objective - DQ may not be business objective

What do I need to take from data governance perspective

David S. not sure where DQ will come

Add bullet and reference SMS COE in this, for example

David S. will get us a guidelines document

They try to align with UHG standards and guidelines

Create a guideline document and add a link to the UHG data quality reference.

Architects will factor it in and ARB will check the implementation of it.

There is report that shows various projects in flight

Identify projects at various stages

Tom J. helps with portfolio intakes

+++++

Meeting 07/07/2023 with Gurtej Singh

Subject: Business metadata, Business metadata in Tableau

DEEP is BDPaaS

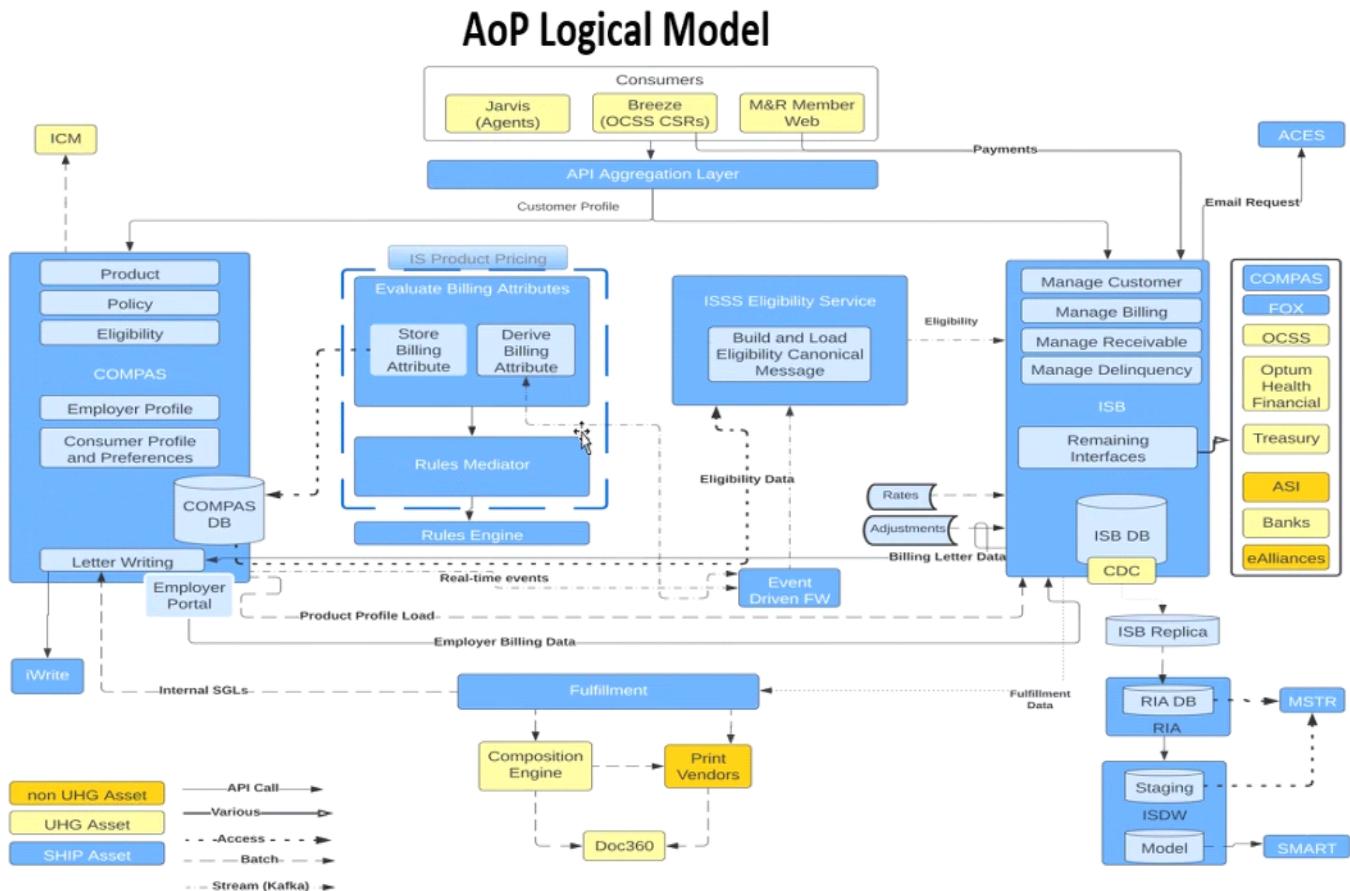
Share the Excel sheet with Gurtej and he will create a metadata dashboard

Doug U. is the right person for future data quality and data governance needs

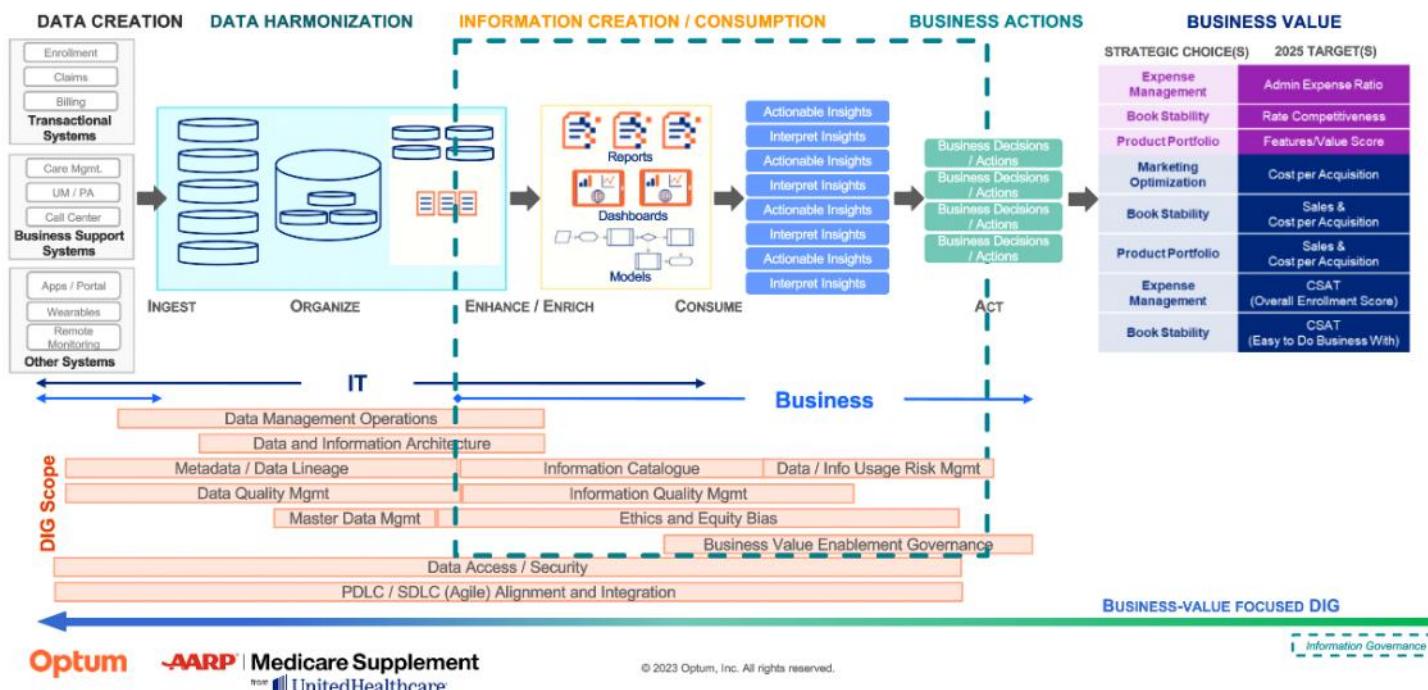
## Data Lineage

Thursday, May 18, 2023 9:44 AM

# Compas -> ISB -> RIA -> ISDW



# Data and information governance across the continuum maximizes business value enablement



# Miscellaneous

Tuesday, May 2, 2023 12:06 PM

[12:01 PM] Cheema, Dave  
Chetan, who can I talk to about how the DQ is defined and implemented at OCDP? Thank you.

[12:02 PM] Koneru, Chetan  
I implemented it. Rules are defined by Param [mailto:paramjit\\_nayak@optum.com](mailto:paramjit_nayak@optum.com)  
He can give you a demo

## MRIS Data Quality

### ASK

Monitor data quality  
How is it implemented?

### Expectations

data quality capabilities and fortify the DQ team to proactively support business needs  
Verify functional fit of Global ID for known DQ needs and help identify alternative technologies  
incorporate DQ into software development and EA processes  
KPI to monitor and measure the business?  
Build technical and business metadata

### DQ Capabilities

Shortlist business areas for DQ issue analysis / capability maturity	Assess data needs, likely impact to business value for each use case	<ul style="list-style-type: none"><li>Business areas for DQ issue analysis / capability maturity<ul style="list-style-type: none"><li>Key value-initiatives (use cases) within each business area?<ul style="list-style-type: none"><li>Categorize new use cases, emerging use cases and mature use cases</li><li>Identify use cases with common data needs across different business areas</li></ul></li></ul></li><li>Lineage of DQ issues to the business process</li><li>Assess data needs, likely impact to business value for each use case?<ul style="list-style-type: none"><li>Define KPI to measure DQ</li><li>DQ metrics and DQ profile published for key datasets</li><li>KPI socialization and cadence</li></ul></li></ul>		
Identify key value-initiatives (use cases) within each business areas	Identify use cases with common data needs across different business areas	<ul style="list-style-type: none"><li>Lineage of DQ issues to the business process</li><li>Assess data needs, likely impact to business value for each use case?</li><li>Define KPI to measure DQ</li><li>DQ metrics and DQ profile published for key datasets</li><li>KPI socialization and cadence</li></ul>		
Categorize new use cases, emerging use cases, mature use cases that are operational	Lineage of DQ issues to the business process	<ul style="list-style-type: none"><li>Business areas for DQ issue analysis / capability maturity<ul style="list-style-type: none"><li>Key value-initiatives (use cases) within each business area?<ul style="list-style-type: none"><li>Categorize new use cases, emerging use cases and mature use cases</li><li>Identify use cases with common data needs across different business areas</li></ul></li></ul></li><li>Lineage of DQ issues to the business process</li><li>Assess data needs, likely impact to business value for each use case?<ul style="list-style-type: none"><li>Define KPI to measure DQ</li><li>DQ metrics and DQ profile published for key datasets</li><li>KPI socialization and cadence</li></ul></li></ul>		
Define KPI to measure DQ	DQ metrics and DQ profile published for key data sets	KPI socialization and cadence		

### DQ Tools

Assess Global ID usage	<ul style="list-style-type: none"><li>Assess Global ID usage?</li><li>Address DQ tools/technology functionality gaps and support technology replacement decision, if needed?</li><li>Metadata integration with DQ?</li></ul>	
Address DQ tools/technology functionality gaps and support technology replacement decision, if needed	<ul style="list-style-type: none"><li>Metadata integration with DQ?</li></ul>	

### SDLC – DQ Adoption

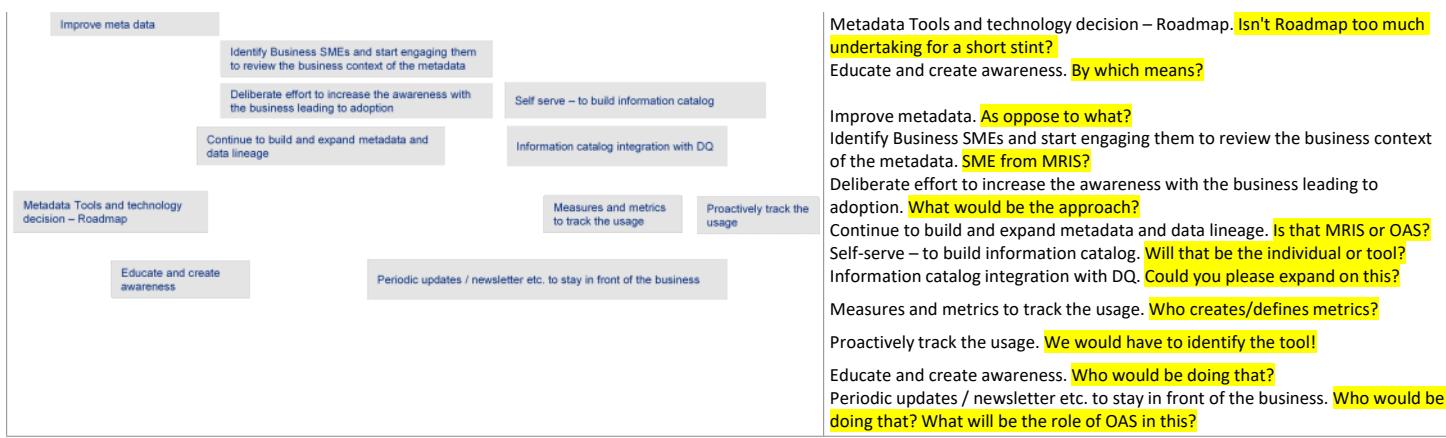
Assess SDLC life cycle across different types of project across different business areas	DQ KPI for SDLC processes	DQ cadence with Data ideation team	DQ Check points into SDLC	<ul style="list-style-type: none"><li>Assess SDLC lifecycle across different project types and business areas. Who identifies and what quantity?</li><li>DQ KPI for SDLC processes. Can you give me an example?</li><li>DQ cadence with data ideation team?</li><li>DQ Check points into SDLC. Can you give me an example?</li></ul>
SDLC process analysis for DQ adoption and integration	Assess EA life cycle			

### Q3 - Q4

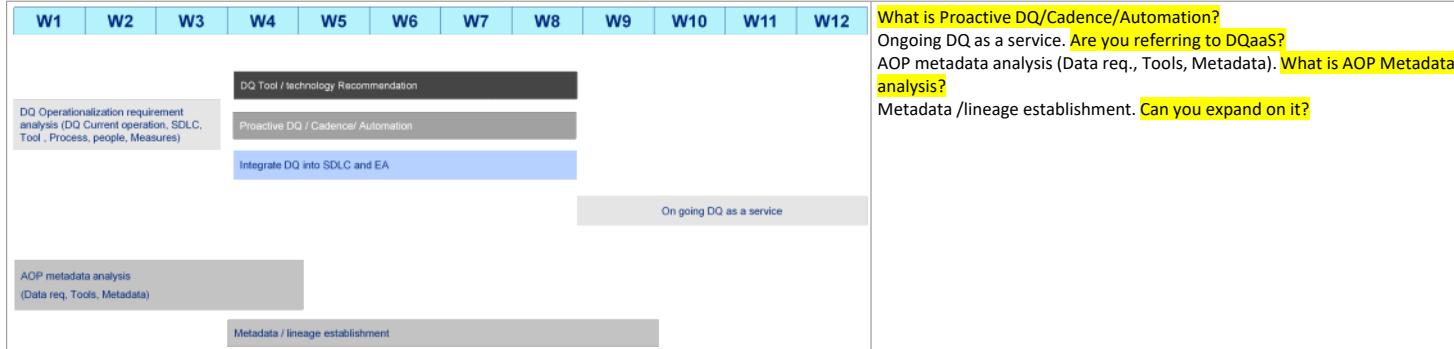
Identify DQ needs across all known use cases (SDLC +), define DQ management strategies (tool, people, process)	Identify action items, Plan and execute Sprints	Integrate DQ with data ingestion and proactive check	Define plan for business adoption to make DQ more proactive and identify opportunities to move it to the left	Automation approach to DQ	<ul style="list-style-type: none"><li>Identify DQ needs across all known use cases (SDLC +), define DQ management strategies (tool, people, process). How is going to provide us this information and what is the timeline for this?</li><li>Identify action items, Plan and execute Sprints. What is the scope here?</li><li>Integrate DQ with data ingestion and proactive check. Explain proactive check?</li><li>Define plan for business adoption to make DQ more proactive and identify opportunities to move it to the left</li><li>Automation approach to DQ?</li></ul>
--	---	--	---	---------------------------	---

### Metadata / Lineage

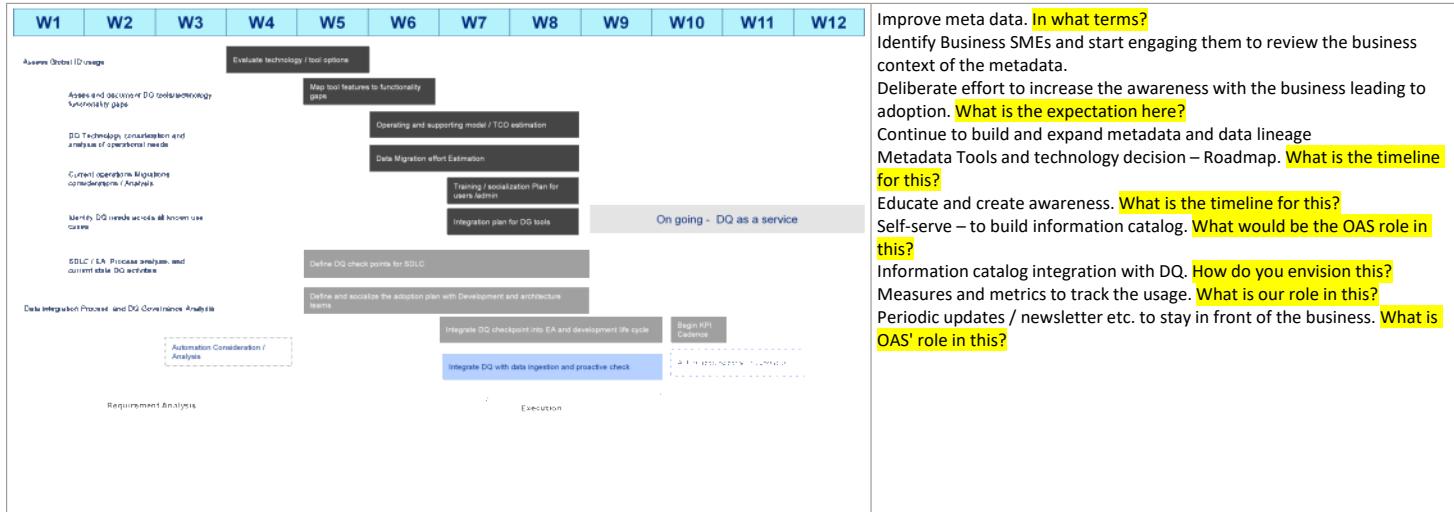
More of a soft change management along with the enriching the repository along with selection for appropriate technology to replace Collibra



#### Proposed – Operationalization of Proactive DQ



#### DQ operationalization and tool evaluation



#### AOP Metadata Analysis

W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12
----	----	----	----	----	----	----	----	----	-----	-----	-----

Data requirement analysis

Metadata / data alignment Analysis\*

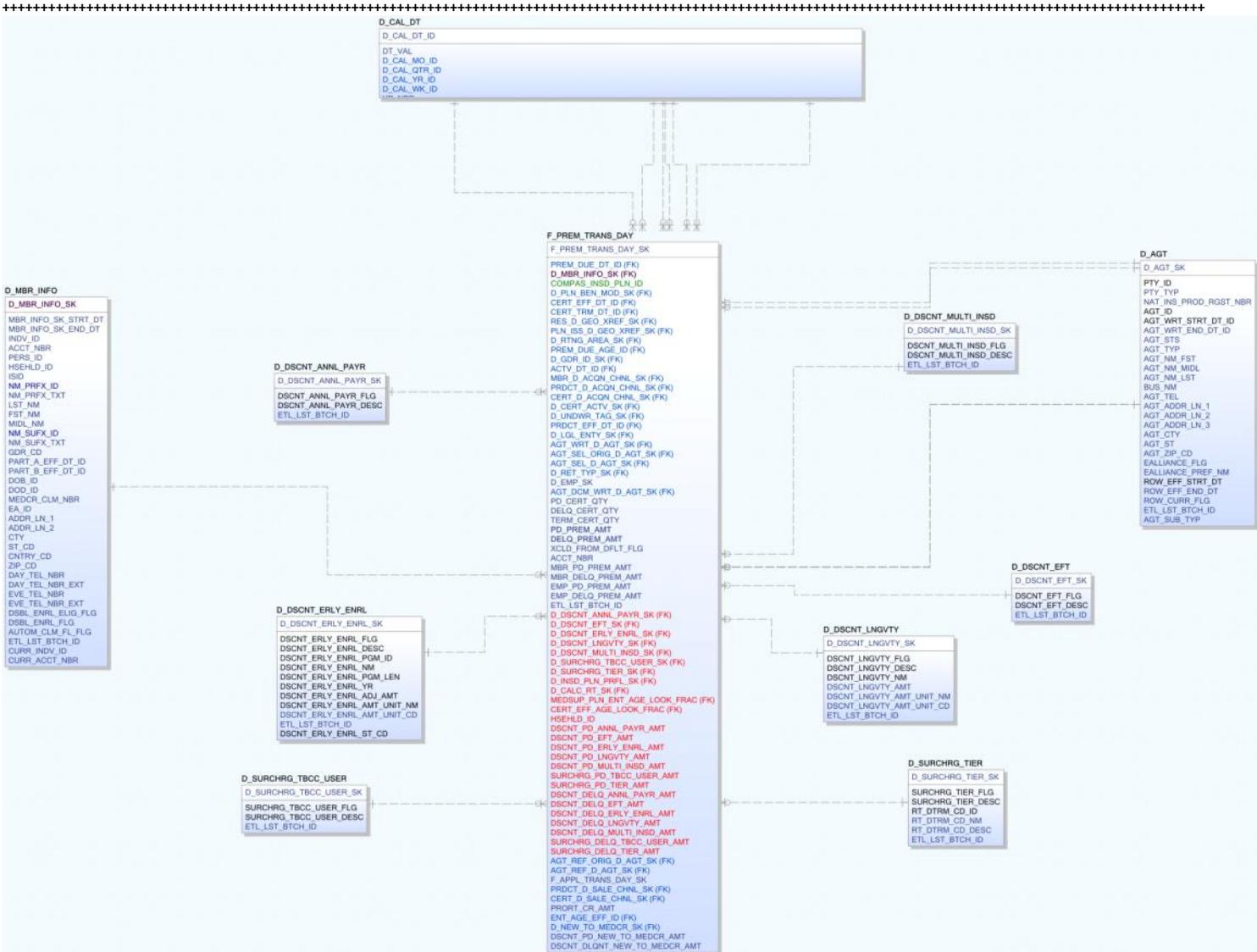
Tool usage / utility analysis

Target Metadata / data mapping analysis \*

Lineage Analysis



\* Completed work product exists from current work and requires only review



### Clarifications 05/11/2023

**May 17 - Processes impacting data elements** - Data elements that are having quality issues and the processes that are causing those quality issues for those data elements.

**May 23 - Operating model, R&R** - who is doing what and what are their roles and responsibilities (R&R) in light of current DQ processes (like intake, how you address, how and who publishes etc.)  
- Learn projected demand - The demand forecast of anticipated DQ issues, if it is adopted properly, terms of volume, velocity, and frequency (mostly the se 2 aspects are to evaluate the right tool )

**May 24 - Data migration/volume/set up estimation** - What is the demand to move data from current DQ tool, will it necessitate any data migration? If yes, what would be the expected volume and will it require, any set up estimates - do we need to work with any other teams ?

- **Automation consideration** - is there any automation, e.g., scripts, currently being used?  
June 1 - **Identification data elements for identified use cases** - We are going to identify some use cases for DQ. This refers to the data elements pertaining to those use cases. - This is the data lineage for all the data elements required to establish the lineage.

---

## Data Quality Tools Evaluation Criteria:

1. Data Profiling and Assessment: The tool should offer comprehensive data profiling capabilities to assess the quality and characteristics of data. It should identify data anomalies, such as duplicates, inconsistencies, and missing values, and provide data quality metrics.
  2. Data Cleansing and Standardization: Look for tools that provide advanced data cleansing and standardization capabilities. The tool should support a wide range of data transformation functions, including parsing, formatting, validation, and enrichment.
  3. Data Integration and Connectivity: Consider the tool's ability to integrate with various data sources, systems, and platforms. It should support multiple data integration methods, such as batch processing, real-time streaming, and API connectivity.
  4. Data Quality Rules and Monitoring: The tool should allow users to define and implement data quality rules and workflows. It should offer monitoring and alerting capabilities to proactively identify and address data quality issues in real-time.
  5. Data Quality Dashboards and Reporting: Evaluate the tool's reporting and visualization capabilities. It should provide intuitive dashboards and reports to track data quality metrics, trends, and issues. Look for customizable reporting options to meet specific business requirements.
  6. Data Governance and Metadata Management: Consider whether the tool supports data governance practices and metadata management. It should provide features for data lineage, data stewardship, data cataloging, and data privacy compliance.
  7. Scalability and Performance: Assess the tool's ability to handle large volumes of data and its performance in terms of processing speed and scalability. It should be capable of processing data at scale without compromising performance.
  8. Machine Learning and Automation: Look for tools that leverage machine learning and automation techniques to improve data quality processes. These features can assist in identifying patterns, suggesting data quality rules, and automating data cleansing tasks.
  9. Ease of Use and User Interface: Evaluate the tool's user interface and ease of use. It should have an intuitive design that enables both technical and non-technical users to navigate and utilize the tool effectively.
  10. Vendor Viability and Support: Consider the vendor's reputation, financial stability, and customer support services. Evaluate their track record in delivering timely updates, addressing customer issues, and providing comprehensive documentation and training.
  11. Cost and ROI: Finally, consider the total cost of ownership, including licensing fees, implementation costs, and ongoing maintenance. Assess the tool's potential return on investment (ROI) by considering the value it can provide in terms of improved data quality, reduced errors, and increased efficiency
- 

## OpenMetadata

### OpenMetadata strengths

OpenMetadata is a powerful tool for metadata management and data governance:

- **Openness:** An open standard, which means that it is freely available to use and modify.
- **Flexibility:** Flexible standard that can be used to support a wide variety of metadata needs. A good choice for organizations that need to manage a variety of different types of data.
- **Reliability:** OpenMetadata is reliable that is developed and tested by a large community of experts.
- **Security:** OpenMetadata provides security to protect sensitive data.
- **Comprehensive Metadata Management:** Provides a comprehensive solution that can capture, organize, and govern metadata across various data sources, including databases, data warehouses, data lakes, and streaming platforms.
- **Centralized Metadata Repository:** Provides a centralized metadata repository that serves as a single source of truth for all metadata related to an organization's data assets.
- **Metadata Exploration and Discovery:** Provides powerful features for metadata exploration and discovery and a unified interface to search and explore metadata.

### OpenMetadata weaknesses

- **Complexity:** A complex tool, and it can be difficult to learn and use.
- **Cost:** Commercial product, and it can be expensive to purchase and maintain.
- **Lack of integration:** Not integrated with many other data management tools.
- **Security:** Not as secure as some other data management tools
- **Scalability:** Can be difficult to scale to meet the needs of large organizations.
- **Customization:** Not as customizable as some other data management tools.
- **Support:** Does not have as much support as compared with some other data management tools.
- **Performance:** May experience performance issues, particularly when dealing with complex queries or large volumes of metadata.
- **Data quality:** Maintaining data quality can be challenging, especially when dealing with diverse data sources, inconsistent formats, or data that is constantly changing.
- **Governance and security:** Ensuring appropriate governance and security measures, such as access control, data privacy, and compliance with regulations, can be a significant challenge for metadata management systems.

Note: OpenMetadata is **not on the PADU list. To put it on the PADU or get an exception can be challenging.**

---

## Product Assessment Framework

A product assessment framework is a structured approach or set of criteria used to evaluate and analyze the various aspects of a product. It helps organizations or individuals make informed decisions about the viability, market potential, and overall quality of a product. Here is a general framework that can be used for assessing products:

1. Market Analysis:
  - Target audience: Identify the specific group of customers the product is intended for.
  - Market size and growth potential: Determine the size of the target market and its potential for growth.
  - Competitive landscape: Analyze the competition and identify the product's unique selling points.
  - Market trends: Consider the current and future trends that may impact the product's success.
2. Value Proposition:
  - Define the unique value the product offers to customers.
  - Assess how the product solves a specific problem or meets a need.
  - Identify the key benefits and advantages of the product over existing alternatives.
3. Product Features and Functionality:
  - Evaluate the core features and functionality of the product.
  - Determine how well the product performs its intended purpose.
  - Assess any unique or innovative features that differentiate it from competitors.
4. User Experience (UX) and Design:
  - Evaluate the overall user experience of the product.
  - Consider factors such as ease of use, intuitiveness, and accessibility.
  - Assess the visual design, aesthetics, and overall user interface.
5. Technical Assessment:
  - Evaluate the technical aspects of the product, such as scalability, reliability, and security.
  - Consider the underlying technology stack and infrastructure.

- Assess any potential limitations or challenges related to implementation or maintenance.
6. Business Model:
    - Evaluate the product's revenue generation potential and profitability.
    - Assess the pricing strategy and business model.
    - Consider factors such as cost structure, revenue streams, and customer acquisition.
  7. Marketing and Distribution:
    - Assess the marketing and distribution strategy for the product.
    - Evaluate how the product will reach its target audience.
    - Consider channels, promotion methods, and partnerships.
  8. Risk Assessment:
    - Identify potential risks and challenges associated with the product.
    - Assess factors such as market competition, regulatory compliance, and intellectual property.
    - Develop mitigation strategies for identified risks.
  9. Financial Analysis:
    - Evaluate the financial viability of the product.
    - Assess revenue projections, cost estimates, and return on investment.
    - Consider factors such as break-even point and payback period.
  10. Feedback and Validation:
    - Gather feedback from potential customers or target audience.
    - Conduct user testing and prototype validation.
    - Incorporate feedback into the assessment process.

It's important to note that the specific criteria and weightage within each category may vary depending on the nature of the product and the goals of the assessment. Organizations may also have their own customized frameworks based on their unique requirements.
  11. Benefits of using a product assessment framework:
    - It can help you to **identify the strengths and weaknesses** of your product.
    - It can help you to **understand your target market and their needs**.
    - It can help you to **identify your competitors and their strengths and weaknesses**.
    - It can help you to **develop a marketing and sales strategy** that will reach your target market.
    - It can help you to **measure the success** of your product and **make necessary adjustments**.
  12. Examples of product assessment frameworks:
    - The Kano Model
    - The Product Opportunity Evaluation Matrix (POEM)
    - The Value Proposition Canvas
    - The Business Model Canvas
    - The Lean Startup Canvas
- 

#### DQ KPIs for SDLC cadence

- Data accuracy: The percentage of data that is accurate.
- Data completeness: The percentage of data that is complete.
- Data consistency: The degree to which data is consistent across different systems and sources.
- Data timeliness: The time it takes to collect, clean, and integrate data.
- Data accessibility: The ease with which data can be accessed and used by stakeholders.
- Data relevance: The percentage of data that is relevant to the business needs.

#### Additional DQ KPIs for SDLC cadence

- Time to detection: time it take detection an issue
  - Time to resolution: time to resolve an issue
  - Escaped Defects: It measures the number of defects that are discovered by customers or end-users after the release or deployment
  - Change Failure Rate: Percentage of failed changes or deployments that result in production issues or incidents.
  - Data Duplication Rate: Duplicate data entries or records throughout the SDLC.
  - Data Integrity: Ensuring data remains reliable, accurate, and consistent.
  - Data Quality Issue Resolution Time: Monitor the time it takes to identify, address, and resolve data quality
  - Data Validation Coverage: Extent to which data validation rules and checks are implemented and executed during the SDLC
  - Code Coverage: Percentage of code covered by automated tests. Higher code coverage, higher code quality.
- 

#### Top data quality tools:

- HomeGrown Tool for quality measurement
  - DQaaS
- Commercial:
  - Informatica Data Quality (IDQ)
  - Talend Data Quality
  - Ataccama Data Quality Management Platform
- Open Source:
  - ~~Talend Open Studio for Data Quality~~ Not on PADU list
  - ~~OpenRefine (formerly Google Refine)~~ Not on PADU list
  - ~~OpenData~~ Not on PADU list
  - ~~Data Wrangler~~ Not on PADU list
  - ~~Trifactor Wrangler~~ On PADU list, however, it is Discouraged

Mike Hastie choice: OpenMetadata (Open-Metadata.org) - Not on PADU list

---

#### Data Lineage Tools

- Commercial
  - Alation Data Catalog
  - Collibra
  - Talend Data Lineage
  - Dremio
  - Informatica Metadata Manager
- Open source
  - Apache Atlas
  - Tokern - is a column-level data lineage tool that can track lineage from source to target, including all transformations and joins.

- Egeria - can collect lineage information from a variety of sources, including databases, ETL tools, and data warehouses.
  - Pachyderm - distributed data science platform. It can track lineage from source to target, including all transformations and joins.
  - OpenLineage - supports a variety of data sources, including databases, ETL tools, and data warehouses. Can collect lineage information automatically or manually.
  - TrueDat - provides a comprehensive view of data movement and transformation. TrueDat can track lineage from source to target, including all transformations and joins.
  - Marquez - provides data lineage and metadata management capabilities.
  - Amundsen primarily discovery and metadata platform, but also provides data lineage functionality.
  - DataHub - provides data lineage tracking, data discovery, and data cataloging capabilities.
  - Metacat Metacat - metadata management and data discovery platform. Also, provides data lineage tracking, data versioning, and other data management capabilities.
  - Dataedo - provides data lineage tracking, data discovery, and data profiling capabilities.
- 

## MRIS DQ Adoption

### ASK

Enhancing and monitoring the data quality (DQ)  
Current DQ process is reactive and it requires to be proactive  
Plan data quality strategically  
OAS to provide continuity support part of DIG assessment

### Our Understanding

Introduce proactive DQ to minimize data quality issues  
Verify functional fit of Global ID and identify alternatives??  
Engage with operations teams to incorporate DQ into software development and EA processes??  
Build KPI's to monitor and measure adoption of DQ??  
Build technical and business metadata. Who identifies the metadata??

### Deliverables

DQ implementation plan for TWO focused data quality issues with go-forward processes for proactive issue identification and remediation  
DQ Tool Recommendations to replace Global ID  
Documented SoPs and implementation plan to incorporate DQ into software development and EA process  
KPI's to monitor and measure the business and technology adoption of DQ for ONE inflight project??  
Metadata lineage for THREE selected use cases for AOP / COMPAS migration. Do we have to implement metadata lineage tool??

### Delivery Assumptions

Project will be an active collaboration between OAS and MRIS with active participation of selected MRIS team members to facilitate active coaching and enablement.

1. Select members of MRIS Data governance and Data architecture team will
  - a. Provide details of current implementation of DQ tool
  - b. Identify and finalize DQ issues that impact business value
  - c. Identify and finalize ONE project for establishing DQ KPI and monitoring. Do we have to setup the environment??
  - d. Identify the person responsible for cadence of KPI
  - e. Identify and finalize the 3 use cases for establishing metadata lineage before the start of the project

### DQ operationalization - Use cases

Here are 2 candidate DQ use cases – one basic profiling case, and one for testing complex business rule:

1. Basic profiling of Agent data – The quality and complexity of Agent data varies. Agents range from large organizations to small individuals, and Agent data is constantly updated with new and different Agents. This makes Agent data a good candidate for regular profiling for basic quality dimensions of completeness, conformity, consistency and integrity. Examples could include:
  - a. Phone numbers –
    - i. how many formats are there? Do phones include dashes or parentheses?
    - ii. Are phone numbers all numeric? Do all phone numbers contain exactly 10 numbers?
    - iii. How often are phone numbers missing?
  - b. State codes –
    - i. Are all State codes found in the State code table?
    - ii. Do zip codes and State codes align?
    - iii. How often are state codes missing?
2. Complex business rule checks for key Member data – member data undergoes many change scenarios and is on many systems. Keeping the member data correct can be challenging. This makes Member data a good candidate for regularly checking accuracy, timeliness and integrity. Examples could include:
  - a. Email address –
    - i. While it is possible for multiple individuals to share an email address (husband and wife under same plan) there are cases where a shared email is incorrect. For example, two different individuals with same name and DOB, but in different insured plans, addresses and households. In other words, two different individuals who share same name and DOB, but are otherwise unrelated.
    - ii. Is email address up to date, and consistent across tables and systems?
    - iii. Is email address always populated and in correct format? Does it contain all lower-case letters, and an "@" sign?
  - b. Multiple location addresses –
    - i. If an individual has multiple addresses for location, are there no more than 2 that are 'active'?
    - ii. If individual has 2 active addresses, is one permanent, and the other temporary?
    - iii. If individual recently moved, have all tables and systems received a timely and correct update?

### Metadata lineage – Use case

#### 3 Use case

- New Plan Add: A new plan got added to an existing Member. The Key attributes that are connected with this change flow from Compas into ISDW
- Smoker Surcharge : Member became a smoker and surcharge is added.
- Member Payment: Member paid his Premium.

Global ID usage assessment??

Any data points collected during client (MRIS) interviews??

In Execution Task Details: Tabulate data elements, process that impact the elements??

Identify the beta projects for DQ operationalization (2 or 3). Are we setting up the environment or we're just defining how to do it??

Tool recommendation - Charter. What does Charter mean in this context??

Automation consideration??

Analysis of development process. But we have not gotten KT from MRIS of the current development process??

Desired Target Feature list for DQ and envisioned Operating model - shouldn't this be MRIS??

Tool recommendation - Charter??

Roadmap / Plan for other ongoing projects. What level of detail, because roadmaps take time??

**Questions:**

Verify functional fit of Global ID and identify alternatives??

Need to understand how they're using it and what are the alternatives

Engage with operations teams to incorporate DQ into software development and EA processes??

Understand how the DQ is integrated into the Architecture and SLDC processes

Build KPIs to monitor and measure adoption of DQ??

Adoption of DQ

Build technical and business metadata. Who identifies the metadata??

Contact Raj, Sreelatha

KPIs to monitor and measure the business and technology adoption of DQ for ONE inflight project??

Identify where toll gates are to be deployed

Any data points collected during client (MRIS) interviews??

Offshore consulting

Tabulate data elements, process that impact the elements??

Tabulation of data elements and processes that impact those elements

Identify the beta projects for DQ operationalization (2 or 3).

Process enhancement for DQ; To identify the project

Tool recommendation - Charter.

Tool recommendation output

Automation consideration??

May be script implementation

Week 3 - get development process and Analysis of development process.

Understand the development process and analyze it to identify DQ improvement areas

Desired Target Feature list for DQ and envisioned Operating model

For the tool only

Roadmap / Plan for other ongoing projects. What level of detail, because roadmaps take time??

Extremely high level set of step to improve data quality

Collateral collected during interviews?

No much available

Tool on-prem or cloud?

Whichever tool we recommend, must be available in cloud

**Use case Analysis**

Identify all the current DQ use cases (agent, complex COMPAS email)

M/O

2

Classify the reactive and proactive use cases

M/O

2

Identify / List business impact of DQ

M/O

2

Tabulate data elements, process that impact the elements

O

Identify data / Process owners (will ask around)

M/O

1

**Data model analysis / Walkthrough**

O

2

**Tool Analysis**

Identify top 3 tools of choice for DQ

M/O

1

Current TCO for DQ / Metadata tool

M

1

Operating model, R & R

M

1

Projected demand

M

1

Data migration needs / volume / Set up estimation

M

1

Automation consideration

O /M

**Shortlist data lineage business Use case (4 ) for AOP**

M /O

2

**Tool for metadata finalization, Template finalization review**

M /O

1

**SDLC Process**

Analysis of development process

O /M

2

Current gaps for DQ / DIG in development process

O/M

2

Analyze Architecture process

O /M

1

Gap identification for DQ / DIG in Architecture process

O /M

1

Identify the beta projects for DQ operationalization (2 or 3)

M

1

Review meetings

**Identification and documentation of all source / target data elements for identified use cases**

M /O

2

**Sample lineage review**

M

1

Use case analysis

Identify DQ use cases

Classify use cases

Identify business impact of DQ

Processes impacting data elements

Identify data / Process owners

**Tool Analysis**

Identify top 3 tools of choice for DQ

Collect current TCO for DQ / Metadata tool

Operating model, R & R

Learn projected demand

Data migration needs / volume / Set up estimation

Automation consideration  
Get data lineage business Use cases (4 ) for AOP  
Tool for metadata finalization, Template finalization review

#### SDLC Process

Learning and analysis of development process  
Gaps identification for DQ in development process

Learn and analyze Architecture process  
Gap identification for DQ in Architecture process

Identify the beta projects for DQ operationalization (2 or 3)  
Identification data elements for identified use cases

Sample lineage review

#### Tool Recommendation and DQ operationalization

Document and socialize SDLC toll gates with SDLC stakeholders  
Document an socialize toll gate process for architecture  
Define and review KPI for SDLC cadence  
Document and review 3 use cases for data lineage to be completed

\*\*\*\*\*  
**Proactive use case** is a scenario in which a **system or process takes action before a problem occurs**. This can be done by monitoring data, identifying potential issues, and preventive steps  
**Reactive use case** is a scenario in which a **system or process responds to changes** in its environment in a timely and appropriate manner. This can be done by monitoring data, identifying changes, and taking steps to adapt.

\*\*\*\*\*  
Pl see this link [ACTIVE FINAL WORKING MRIS DIG Current State Report\\_client Sharable.pdf](#)  
[ACTIVE FINAL WORKING MRIS DIG Current State Report\\_client Sharable.pdf](#)

### Steps to improve data quality

- Define Data Quality Objectives:** Clearly identify the desired level of data quality based on your business requirements. Determine specific quality metrics, such as accuracy, completeness, consistency, and timeliness.
- Assess Current Data Quality:** Evaluate the existing data quality by conducting a thorough assessment. Identify data issues, inconsistencies, and areas for improvement. Use data profiling techniques, data sampling, and data quality tools to analyze the quality of your data.
- Establish Data Quality Standards:** Define data quality standards and guidelines that align with your objectives. These standards should encompass data formats, naming conventions, data validation rules, and business rules.
- Implement Data Validation Processes:** Develop data validation processes to ensure data integrity and accuracy. This includes performing checks for data completeness, consistency, uniqueness, and referential integrity. Use automated validation techniques and tools to streamline the process.
- Cleanse and Standardize Data:** Apply data cleansing techniques to rectify errors, remove duplicates, and handle inconsistencies. Utilize tools and techniques for data deduplication, standardization, and normalization to improve the quality and consistency of your data.
- Data Integration:** Ensure seamless integration of data from various sources. Implement robust data integration processes to consolidate data, resolve conflicts, and eliminate data silos.
- Data Security and Privacy:** Implement robust security measures and adhere to data privacy regulations to protect sensitive information. Ensuring data confidentiality, integrity, and availability contributes to data quality.
- Establish Data Governance:** Implement a data governance framework to ensure accountability, ownership, and responsibility for data quality. Establish data stewardship roles and define data governance processes to enforce data quality standards.
- Implement Data Quality Controls:** Establish data quality controls to monitor data quality on an ongoing basis. Set up automated monitoring mechanisms, data quality dashboards, and exception reporting to detect and address data issues promptly.
- Conduct Data Quality Audits:** Regularly perform data quality audits to assess the effectiveness of your data quality processes. Identify areas of improvement, validate adherence to standards, and ensure continuous enhancement.
- Enhance Data Documentation:** Maintain comprehensive documentation of your data sources, definitions, transformations, and business rules. Document data lineage and metadata to ensure transparency and facilitate data understanding.
- Train and Educate Users:** Provide training and education to individuals responsible for managing and using data. Educate them about data quality concepts, best practices, and their role in maintaining data quality.
- Foster a Data Quality Culture:** Cultivate a data-driven culture within your organization that prioritizes data quality. Encourage collaboration, communication, and awareness around the importance of data quality across all levels.
- Continuously Improve:** Data quality is an iterative process. Continuously evaluate and refine your data quality practices based on feedback, emerging technologies, and changing business needs. Regularly review and update data quality standards and processes.

### Data Quality & Observability: DQ Rule Cheat Sheet

Which DQ Problem Do You Have?

Data Quality & Observability auto-discovers issues in data using a machine learning first, rules second based approach. It uses associative, unsupervised machine learning to auto generate SQL- based, explainable and adaptive data quality rules. It creates snapshots and baselines to benchmark past data, constantly learns from new data and makes predictions for typos, formatting issues, outliers, relationships and more.

1. <b>Datasypes:</b> How many phone number formats are in this column?	Shape, Format, Conformity	Common in STRING fields or fields defined as VARCHAR you can end up with many different formats that express something like a zip code or phone number or SSN for example. It is helpful to find the majority formats and show the topN data shapes that make up the column values. This helps identify typos and strange formats. Quick demo video: <a href="#">PII &amp; Data Quality</a>
2. <b>Row count check:</b> Has my row count dropped on any dataset?	Behavior	It can be important to know if the volume of a dataset drops, also known as a row count drop. When a dataset suddenly has fewer rows than normal it can mean data is missing in the file or table
3. <b>Business DQ rules:</b> I need to write custom DQ mics.	Custom rules	Define rules in standard SQL. There are about 105 predefined functions included but not limited to concat, length, avg, max, min, substr etc. Most common use-cases are easily covered by vanilla SQL but there are more advanced situations where window functions, grouping and time-series functions are required. These can also be expressed fairly easily with some Collibra DQ SQL extensions. Quick demo video: <a href="#">Quickly add rules</a>
4. <b>Auto-generated &amp; adaptive DQ rules:</b> I want DQ rules generated for me.	Adaptive rules	Out of the box unless turned off, Collibra DQ will start generating many rules for each column based on its profile over time. It compares each daily run with its baseline. It helps you avoid writing and managing thousands of conditional statements. Leverage ML to auto-generate SQL-based, non-proprietary, explainable and adaptive data quality rules. Reduce manual rule management efforts. Quick demo video: <a href="#">Out-of-the-box features</a>
5. <b>Fuzzy matching:</b> Is Bill Gates the same as William Gates in my database?	Dupe or Uniqueness	Commonly we need to find exact duplicates or similar duplicates. This problem is not suited for conditional statements. Collibra DQ allows you to opt into any grouping of columns and find exact or similar records, ie fuzzy matching. This can be done at the column or record level. Identify duplicate or redundant data across multiple data domains via fuzzy or exact matching. Quick demo video: <a href="#">Duplicates or Fuzzy Matching</a>
6. <b>NULL check:</b> Do I have complete observations in a world of fill-out forms and manual data collection?	Completeness	Out of the box every column will have a NullCheck in place, the null check is generated from the columns' past behavior or descriptive statistics. You can turn it off per dataset or per column or add manually if desired
7. <b>Valid values:</b> A valid FICO credit score is between 300 and 850.	Validity	One of the most common DQ rules is a valid range of values or nickname valid values. In the case of a credit_score a rule such as where credit_score between 300 and 850. Another example for string values could be where credit_providers IN ('experian', 'trans union', 'equifax'). Both Numeric and String values can be

		expressed as valid values.
8. <b>Change detection:</b> Tell me when something suddenly changes in my data.	Behavior (data drift)	Any column, schema or cell value that suddenly breaks its past trend. Would require thousands of conditional statements and their ongoing management. By default Collibra DQ behavioral analytics is turned on for automatic change control. Quick demo video: <a href="#">Data Drift &amp; Shift Detection</a>
9. <b>PII or sensitive data discovery:</b> Do I have sensitive data in my datasets?		Ability to filter and select datasets by PII, MNPI, PHI etc. During the daily profiling Collibra DQ discovers and tags sensitive data (i.e. SSN). The software keeps an index of all columns that contain sensitive data for interactive filtering from the UI. Automatically understand the semantic schema of your data to classify and mark sensitive data. Quick demo video: <a href="#">Discover with DQ Rule Enforcement</a>
10. <b>Run DQ on files:</b> I need to be able to run DQ rules on files.		Many DQ frameworks do not cover files. Collibra DQ operates on files with equal parity as database tables. Any rule that works on a table will also work the same on file.
11. <b>Real-time DQ analysis:</b> I need to run DQ rules on streaming and sensor data (Kafka).		My data does not exist in a table or file but in a Kafka Topic. I want to detect realtime and unsupervised anomalies in streaming data (constantly flowing messages, jsons, avro or batch data) and sensor data (a standard time-series with signal, time and value). Quick demo video <a href="#">C&lt;?lit?ra Data Quality Kafka</a>
12. <b>Outliers:</b> I need to detect outliers per grouping.		Sometimes, basic column level outliers do not solve the issue. One needs to create a subgroup like stock "symbols" and find outliers over time across all symbols. "I don't care that I have penny stocks and stocks that trade over \$200,000 a share, I care if any one particular stock breaks its past trend or baseline." This is applied when a user wants to find egregious numeric values relative to the population. This identifies invalid entries that are numerically out of place. This also allows for grouping on sub-dimensions to identify anomalies on subsets of data Quick demo video: <a href="#">Outlier Detection</a>
13. <b>Conditional DQ rules:</b> Do I have any rows with dates before 1970?	Conditional rule	Sometimes we just need to express a simple rule. In this example we would click add rule and type d_date < 1970. Any record that contains a d_date before 1970 will produce a break record.
14. <b>Complex DQ rules:</b> I need to join two datasets before I can write a rule.	Complex rule	It is a common need to reference another dataset as a lookup or validation point. Loan credit rates might be stored in a table updated each minute with new loan rates while loan applications might come in one by one or in batches. Quick demo video: <a href="#">DQ Rule Builder Join Example</a>
15. <b>Pipeline testing:</b> I need DQ in my data pipeline.	DQ pipeline	I already have a data pipeline in python or scala or spark and want to control the DQ operations. Some call this an ETL pipeline, making this ETLQ. Collibra DQ supports inline spark commands or REST API commands depending on the need.
16. <b>Distribution rule:</b> Am I missing data for a subset of my dataset?		Track valid values as well as valid distribution of values Equifax -> 400   TransUnion -> 300   Experian -> 7. Quick demo video: <a href="#">Shape Detection Distribution Rule</a>
17. <b>Validate source:</b> I need to compare two tables	Source - Validating source to target accuracy	It is common to need validation when loading data from a file into a database table or from a source database into a target database to identify missing records, values and broken relationships across tables or systems. Quick demo video: <a href="#">Cross Table Validations</a>
18. <b>Pulse View:</b> I'd like to see a heatmap of where all my data errors exist.	Pulse view	Visualize a blindspot heatmap by time, business units and scheduled jobs
19. <b>Cross-column anomalies:</b> The state and zip code don't belong to each other in my dataset	Patterns	Define relationships not rules for sophisticated use cases. This is used for identifying cross-column anomalies. Commonly used for hierarchical and parent/child mismatches. Quick demo video: <a href="#">Misclassified Data Cross Column Anomalies</a>
20. <b>Schema drift detection:</b> Is there any change in my database structure?	Schema columns added or dropped	Fields, columns, and types can be added, removed, or changed

DQ Dimension	Collibra DQ Sub Type	Collibra DQ Type/Class	Collibra DQ Feature	Examples
Completeness	NULL EMPTY	PROFILE BEHAVIOR RULES OUTLIER PATTERN SOURCE RECORD SCHEMA SHAPES	100% per column is 100% complete. All missing sub percentages will report % complete against the dataset.	Fname Kirk [NULL] Brian
Uniqueness	Exact match Fuzzy match	PROFILE BEHAVIOR RULES SOURCE RECORD DUPES	Define exact match or fuzzy match or identify aPK. Collibra DQ Dupe Feature	Kirk haslbeck Kirk hasslbeck
Validity	Format Data shift Cardinality (Distribution shift) Type shift	RULES OUTLIER PATTERN SOURCE SHAPES	Collibra DQ unsupervised learning detects infrequent shapes and data type shifts.	OCT-20 2020-10-01
Accuracy	Numeric Categorical	PROFILE BEHAVIOR RULES OUTLIER PATTERN SOURCE RECORD SCHEMA DUPES SHAPES	Collibra DQ anomaly engine detects values outside of normal or expected range	Sym, price Goog, 1,535.0 Goog, 15 35
Timeliness	Load time	BEHAVIOR RULES SOURCE RECORD SCHEMA	Collibra DQ Behavior Feature	8:01pm, 8:02pm, 9:30pm
C onsistency	Rows, schema, cell_values	RULES OUTLIER PATTERN SOURCE SHAPES	Collibra DQ Validate Source Feature	'USA' -> 'USA'

---



#### PADU acceptable products:

Informatica Data Quality (IDQ)  
Talend Data Quality

[11:06 AM] Cheema, Dave

Sojan, what is the underlying technology where we'd be applying MRIS DQ, e.g., Hive, Snowflake, Data Lake, Kafka, etc.? Thanks in advance.

[11:09 AM] Paul, Sojan K

Our back-end is pretty much Exadata/Oracle. We are looking into migrating our Data Warehouse to Snowflake later this year/early 2024 but our transnational systems will stay on Exadata at least for another year

[12:37 PM] Cheema, Dave

Sojan, we're in the process of evaluating DQ tools for the MRIS DQ project and a question about the open source tool/product came up. I have the following questions: 1. are open source products allowed; 2. Are there any rules/qualifications that need to be followed? If yes, what are those rules? Thank you.

[12:41 PM] Paul, Sojan K  
 Open source tools/products are ok as long as it's supported by Optum. We like to use the tools supported by Optum as a service. This way, we can avoid the overhead of managing our own servers/db etc. and maintenance like patching etc.

## Actuarial - to compile and analyze statistics and use them to calculate insurance risks and premiums

Criteria List #1	Criteria List #2	Criteria List #3
<p><b>1. Data Profiling and Assessment:</b> The tool should offer comprehensive data profiling capabilities to assess the quality and characteristics of data. It should identify data anomalies, such as duplicates, inconsistencies, and missing values, and provide data quality metrics.</p> <p><b>2. Data Cleansing and Standardization:</b> Look for tools that provide advanced data cleansing and standardization capabilities. The tool should support a wide range of data transformation functions, including parsing, formatting, validation, and enrichment.</p> <p><b>3. Data Integration and Connectivity:</b> Consider the tool's ability to integrate with various data sources, systems, and platforms. It should support multiple data integration methods, such as batch processing, real-time streaming, and API connectivity.</p> <p><b>4. Data Quality Rules and Monitoring:</b> The tool should allow users to define and implement data quality rules and workflows. It should offer monitoring and alerting capabilities to proactively identify and address data quality issues in real-time.</p> <p><b>5. Data Quality Dashboards and Reporting:</b> Evaluate the tool's reporting and visualization capabilities. It should provide intuitive dashboards and reports to track data quality metrics, trends, and issues. Look for customizable reporting options to meet specific business requirements.</p> <p><b>6. Data Governance and Metadata Management:</b> Consider whether the tool supports data governance practices and metadata management. It should provide features for data lineage, data stewardship, data cataloging, and data privacy compliance.</p> <p><b>7. Scalability and Performance:</b> Assess the tool's ability to handle large volumes of data and its performance in terms of processing speed and scalability. It should be capable of processing data at scale without compromising performance.</p> <p><b>8. Machine Learning and Automation:</b> Look for tools that leverage machine learning and automation techniques to improve data quality processes. These features can assist in identifying patterns, suggesting data quality rules, and automating data cleansing tasks.</p> <p><b>9. Ease of Use and User Interface:</b> Evaluate the tool's user interface and ease of use. It should have an intuitive design that enables both technical and non-technical users to navigate and utilize the tool effectively.</p> <p><b>10. Vendor Viability and Support:</b> Consider the vendor's reputation, financial stability, and customer support services. Evaluate their track record in delivering timely updates, addressing customer issues, and providing comprehensive documentation and training.</p> <p><b>11. Cost and ROI:</b> Finally, consider the total cost of ownership, including licensing fees, implementation costs, and ongoing maintenance. Assess the tool's potential return on investment (ROI) by considering the value it can provide in terms of improved data quality, reduced errors, and increased efficiency</p>	<p>1. Connectivity: Access and apply data quality across a wide range of data sources, including internal/external, at rest/streaming, on-premises/cloud, relational/nonrelational data sources.</p> <p>2. Profiling: The statistical analysis of diverse datasets including structured/unstructured and on-premises/cloud datasets to give business users insight into the quality of data and enable them to identify data quality issues.</p> <p>3. Analytics and interactive visualization: Interactive analytical workflow and visual output of statistical analysis to help business and IT users identify, understand and monitor data quality issues and discover patterns and trends over time, such as through reports, scorecards, dashboards and mobile devices.</p> <p>4. Monitoring and detection: Perform data quality monitoring based on preconfigured, custom-built monitoring rules, or adaptive rules, and alert violations. Automatically detect outliers, anomalies, patterns and drifts. Also provide monitoring dashboard, log files or audit trail for compliance requirements.</p> <p>5. Parsing, standardizing and cleansing: The decomposition and formatting of diverse datasets based on government, industry or local standards, business rules, knowledge bases, metadata, and machine learning (ML). Modification of data values to comply with domain restrictions, integrity constraints or other business rules.</p> <p>6. Matching, linking and merging: Match, link and merge related data entries within or across diverse datasets using a variety of traditional and new approaches such as rules, algorithms, metadata, artificial intelligence (AI) and ML.</p> <p>7. Metadata and lineage: Collect and discover metadata or import metadata from third-party tools. Build or import lineage to perform rapid root cause analysis of data quality issues and impact analysis of remediation. Apply active metadata findings, make use of metadata-based rule recommendations and associations, data discovery and cataloging, metrics view against critical data elements.</p> <p>8. Multidomain support: Address multiple data subject areas (such as various master data domains and vertical industry domains) and offer depth of packaged support (such as prebuilt data quality rules) for these subject areas.</p> <p>9. Address validation/geocoding: Support location-related data standardization and cleansing. Completion for partial data in real-time or batch process.</p> <p>10. Data curation and enrichment: Integrate externally sourced data to improve completeness and add value.</p> <p>11. Rule management and data validation: Design, create and deploy business rules for specific data values. The rules can be called within the solution or by third-party applications for data validation purposes, which can be done in batch or real-time mode.</p> <p>12. Business-driven workflow and issue resolution: The processes and user interface to manage the data quality issue resolution through the stewardship workflow. Easily identify, quarantine, assign, escalate and resolve data quality issues, facilitated by collaboration, pervasive monitoring and case management.</p> <p>13. DataOps support: Collaboration of data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization.</p> <p>14. Deployment environment: Styles of deployment, hardware, operating system and maintenance options for deploying data quality operations.</p> <p>15. Architecture and integration: Commonality, consistency, and interoperability among various components of the data quality toolset (including third-party tools) and other data management solutions or components.</p> <p>16. Automation and augmentation: Automate data quality processes or perform in-depth and rapid execution of data quality actions on datasets by leveraging advanced technologies such as AI/ML, knowledge graph, active metadata or NLP to minimize manual effort.</p> <p>17. Multipersona usability: Suitability of the solution to engage and support the various roles (especially nontechnical</p>	<p><b>1. Product Features and Functionality</b></p> <ol style="list-style-type: none"> <li>Evaluate the core features and functionality of the product</li> <li>Determine how well the product performs its intended purpose</li> <li>Assess any unique or innovative features that differentiate it from competitors</li> </ol> <p><b>1. User Experience (UX) and Design</b></p> <ol style="list-style-type: none"> <li>Evaluate the overall user experience of the product</li> <li>Consider factors such as ease of use, intuitiveness, and accessibility</li> <li>Assess the visual design, aesthetics, and overall user interface</li> </ol> <p><b>1. Technical Assessment</b></p> <ol style="list-style-type: none"> <li>Evaluate the technical aspects of the product, such as scalability, reliability, and security</li> <li>Consider the underlying technology stack and infrastructure</li> <li>Assess any potential limitations or challenges related to implementation or maintenance</li> </ol> <p><b>1. Value Proposition</b></p> <ol style="list-style-type: none"> <li>Define the unique value the product offers to customers.</li> <li>Assess how the product solves a specific problem or meets a need.</li> <li>Identify the key benefits and advantages of the product over existing alternatives.</li> </ol> <p><b>1. Financial Analysis</b></p> <ol style="list-style-type: none"> <li>Evaluate the financial viability of the product.</li> <li>Assess revenue projections, cost estimates, and return on investment.</li> <li>Consider factors such as break-even point and payback period.</li> </ol> <p><b>1. Risk Assessment</b></p> <ol style="list-style-type: none"> <li>Identify potential risks and challenges associated with the product</li> <li>Assess factors such as market competition, regulatory compliance, and intellectual property.</li> <li>Develop mitigation strategies for identified risks.</li> </ol> <p><b>1. Feedback and Validation</b></p> <ol style="list-style-type: none"> <li>Gather feedback from potential customers or target audience.</li> <li>Conduct user testing and prototype validation.</li> <li>Incorporate feedback into the assessment process.</li> </ol> <p><b>8. Benefits of using a product assessment framework</b></p> <ol style="list-style-type: none"> <li>It can help you to identify</li> </ol>

		business roles) required in a data quality initiative: data engineers, stewards, DQ analysts, data architects, DI analysts, business analysts, data preparation and so on	
			the strengths and weaknesses of your product.
			ii. It can help you to understand your target market and their needs.
			iii. It can help you to identify your competitors and their strengths and weaknesses.
			iv. It can help you to develop a marketing and sales strategy that will reach your target market.
			v. It can help you to measure the success of your product and make necessary adjustments.



#### People/meetings:

- o Pam Hoffman for homegrown MDM quality validation
- o FOX contact for FOX pre-batch quality checks
- o Verma for SMART source data quality checks
- o Sheharyar for EDP
- o Gurtej/Doug for DEEP
- o Doug for ISDW
- o Teena from Fox QA
- o Jenn from Fox
- o Adam Cimino for COMPAS
- o Someone for AOP

#### General items:

- What is data quality mean in big data
- What does it mean to "profile" big data
- Expand thought from RDMS to other data types – KAFKA, MONGO, HADOOP, etc

Process	Owner
MDM Quality Validation	Pam Hoffman
Fox Pre-batch quality checks	Fox technical team
SMART source data quality check	Verma
EDP	Sheharyar
DEEP	Gurtej, Doug U.
ISDW	Doug U.
Fox QA	Teena, Jenn
COMPAS	Adam Cimino
AOP	??

#### Q. What does data quality mean in big data, what does it mean to "profile" big data?

Data quality refers to the accuracy, completeness, consistency, relevancy, reliability, and timeliness of the data. Data Quality can affect the accuracy and reliability of the results of big data analytics.

Data profiling is a process that involves examining and analyzing data to gain an understanding of its structure, content, and quality. It creates a summary or profile of the data, providing insights into its characteristics and identifying potential issues or anomalies. Data profiling helps understand the data's distribution, patterns, relationships, and data quality issues.

Techniques such as statistical analysis, data visualization, and exploratory data analysis can be used for data profiling. Data attributes such as data types, ranges, frequencies, uniqueness, and completeness, and data profiling help in uncovering data quality issues such as missing values, duplicates, inconsistencies, or outliers.

By profiling big data, organizations can identify and address potential data quality issues to ensure the results of big data analytics are accurate and reliable.

#### Q. What is full list of data technologies we use? I know we have DB2, Oracle, Hive, Kafka, Hadoop, MongoDB

Oracle, DB2, SQL Server, BDPAAS, MongoDB, Hadoop, MicroStrategy, Exadata, SAS, Tableau, BDPAAS, HDFS, Hive, Pig, Spark, SQoop, Kafka, MDP DW, SFTP, Snowflake, Structured/semi-structured/unstructured data sources, ODBC/JDBC, Linux, Salesforce, R, IBM AIX, Trillium, Informatica, Informatics workflows, DM Express, SyncSort, Korn shell scripts, PowerExchange, Unica(RHEL), Nexedia (voice)

**Q. What is list of data technologies covered by DQaaS?**

Hive, Oracle, SQL Server, DB2, Teradata, Snowflake, MySQL, PostgreSQL, File Based (file formats: Avro, Parquet, delimited files), Kafka

Pre-built data source connectors such as - Kafka, file formats (e.g. Avro, Parquet, delimited files)

JDBC, REST, GQL, DQ compute runtime, Declarative YAML jobs, DQ Central Repository, Canned Dashboards

---

**Q. What is difference between profiling data that is on-premise vs. data that is stored in the Cloud?**

There are several key differences between profiling data on-premises and in the cloud.

Category	Cloud	On-Premise
Cost	Cloud-based data profiling solutions are typically more cost-effective than on-premises solutions, as you only pay for the resources you use.	On-premises solutions require you to purchase and maintain hardware and software, which can be expensive.
Scalability	Cloud-based data profiling solutions are more scalable than on-premises solutions, since you can easily add or remove resources as needed.	On-premises solutions can be more difficult to scale, as you may need to purchase additional hardware or software.
Ease of use	Cloud-based data profiling solutions are typically easier to use than on-premises solutions, as they are designed to be user-friendly and accessible from anywhere.	On-premises solutions can be more complex to use as they may require specialized knowledge and skills.
Security	Cloud-based data profiling solutions are typically more secure than on-premises solutions, since they are hosted in secure data centers.	On-premises solutions may be more vulnerable to security threats, as they are located in your own environment.
Size	If you have a large amount of data, then a cloud-based solution may be a better option, as it can scale more easily.	Due to the fixed infrastructure nature of on-prem solutions may not be ideal for large data sizes.
Data Complexity	Since new tools are being built and evolving in the cloud, for complex data, cloud-based solution may not be the best choice.	For complex data, you may need a solution that offers a wider range of tools and algorithms. In this case, an on-premises solution may be a better choice.
Advantages	<ul style="list-style-type: none"> <li>• More cost-effective</li> <li>• Easier to scale</li> <li>• Easier to use</li> <li>• Secure</li> </ul>	<ul style="list-style-type: none"> <li>• Greater control over the data and the environment</li> <li>• Ability to use custom tools and algorithms</li> <li>• More secure</li> </ul>
Disadvantages	<ul style="list-style-type: none"> <li>• Less control over the data and the environment</li> <li>• Limited choice of tools and algorithms</li> <li>• Less secure</li> </ul>	<ul style="list-style-type: none"> <li>• More expensive</li> <li>• More difficult to scale</li> <li>• More complex to use</li> </ul>
Overall	Cloud-based data profiling solutions are cost-effective, scalable, and easy-to-use.	On-premises solutions provide highly secure and can be hosted in your own environment.

---

+

## Correlation between clearly understood requirements and data quality

There is a correlation between clearly understood requirements and data quality, as the quality of data is greatly influenced by the clarity and completeness of the requirements. Let's explore this correlation in more detail:

- Accurate Data Collection:** Clearly defined requirements help establish what data needs to be collected. When requirements are well-understood, data collection processes can be designed accordingly, ensuring that the necessary data is captured accurately.
- Relevant Data:** Clear requirements also help identify the specific data elements that are relevant to the problem at hand. By focusing on the essential data, unnecessary or irrelevant data can be excluded.
- Consistency and Standardization:** Well-understood requirements enable standardization of data collection and storage processes. Consistent data formats, naming conventions, and validation rules can be established based on clear requirements.
- Data Completeness:** Clear requirements help define the expected level of data completeness.
- Data Validation:** Clear requirements provide a basis for data validation. Validation rules and checks can be defined based on the requirements to verify the accuracy, integrity, and consistency of the data.
- Effective Data Governance:** Clearly understood requirements contribute to effective data governance practices. Data governance involves managing data quality, integrity, security, and compliance.

In summary, clearly understood requirements provide the foundation for effective data management practices. They ensure that the right data is collected, processed, and validated, leading to improved data quality.

---

## How to improve requirements collection?

Improving requirements collection is essential for the success of any project. Here are some tips to enhance the process:

- Engage stakeholders:** Involve all relevant stakeholders in the requirements collection process, including end-users, clients, managers, subject matter experts, and other individuals who will be affected by the project.
  - Clearly define project goals:** Before initiating requirements collection, establish a clear understanding of the project's objectives and desired outcomes. This helps in aligning the requirements with the overall vision and ensures that collected information is relevant and focused.
  - Use multiple techniques:** Leverage a variety of requirements gathering techniques such as, interviews, surveys, workshops, observations, and document analysis, to gather a comprehensive set of information.
  - Ask open-ended questions:** Ask open-ended questions that encourage stakeholders to provide detailed responses.
  - Prioritize requirements:** Prioritize requirements based on their importance, urgency, and feasibility to focus on managing scope and focus on high-priority items.
  - Ensure clarity and specificity:** Document requirements in a clear and specific manner. Avoid ambiguity, vagueness, or overly technical language.
  - Validate and verify requirements:** Regularly validate and verify requirements with stakeholders to ensure accuracy and completeness using reviews, walkthroughs, and prototyping sessions.
  - Document requirements systematically:** Adopt a structured documentation process to capture and organize requirements effectively. Usage of templates, diagrams, and other tools to document requirements in a consistent manner, making them easier to understand, review, and track throughout the project life cycle.
  - Manage changes effectively:** Requirements may evolve throughout the project. Establish a robust change management process that allows for the controlled modification of requirements.
  - Collaborate and communicate:** Encourage collaboration and open communication channels among project team members, stakeholders, and the requirements collection team.
- 

## Data Quality / Usage Impact

- Data Availability
- Timeliness
- Accessibility
- Consistency
- Comprehension
- Accuracy
- Completeness
- Trustworthiness
- Data Sharing

Your status is set to do more chats. You'll only get notifications for urgent messages and from your priority contacts. Change settings

**Teams**

Recent

- General  
Insurance Solutions - DIG
- Your items
- Hadoop Data Platform User Group
- General
- Canary
- Change Data Capture
- Collaboration Migration Support
- Data Catalog Support
- Data Warehouse as a Service
- Ellipsis
- Ellipsis - Test Harness
- Elasticsearch
- Kafka
- Pull Request (MySql, Cens, ED)
- Standard Streams
- Tokenization
- Hidden channels
- Optymize PAC
- General
- Q4H Data estate initiative
- SAS Migration to Open-source
- Insurance Solutions - DIG
- General
- PTS Commercial Growth Enablement
- QAS\_HCP\_Risk\_Data\_Assessment
- QAS - Player Tech Ops - Sales
- QAH Revenue Cycle Project
- Cloud and Data Led Transformation C...
- General
- Assets and Accelerators

Insurance Solutions - DIG ...

Information Solutions - DIG

Members Channels Analytics Apps Tags

Search for members

Add member

Owners (1)

Name	Title	Location	Tags	Role
Nand, Durga	Sr Director, Advisory Srvcs	M983-1000		Owner

Members and guests (8)

Name	Title	Location	Tags	Role
Boettcher, Jacob T.	Advisory Svcs Analyst	M1101-8800		Member
Mahendra, Dinesh	VP, Advisory Srvcs	L889-1000		Member
Shiv, John S.	Director, Advisory Srvcs	M1880-1000		Member
Kanakishnan, Anantha	Sr Director, Advisory Srvcs	M1990-1000		Member
Mandal, Aditya	Advisory Svcs Sr Consultant	I840-1000		Member
Cheema, Dave	Director, Advisory Svcs	C4999-1000		Member
Dubrov, Eric B.	Advisory Svcs Sr Consultant	D1081-1000		Member
Meena, Nishi K.	Advisory Svcs Assoc Analyst	I848-1000		Member

Ramakrishnan, Anantha

11:58 AM

## DQ Maturity model

## Ad-hoc

## Reactive

## Proactive

## Integrated

Opti

- or

  1. Initial Stage: In this stage, data quality practices are ad hoc and reactive. There is little awareness or formalized approach to managing data quality. Organizations at this level often experience data issues and errors without a systematic process to identify, track, and resolve them.
  2. Repeatable Stage: At this stage, organizations begin to recognize the importance of data quality and establish some basic processes and guidelines. Data quality is addressed on a project-by-project basis, and there may be some rudimentary data profiling or validation activities in place. However, these practices are not consistently applied across the organization.
  3. Defined Stage: In the defined stage, organizations have defined and documented data quality standards, policies, and procedures. There is a centralized data quality team or function responsible for monitoring and improving data quality across the organization. Data quality metrics and measurements are established, and data issues are tracked and reported regularly.
  4. Managed Stage: At this stage, data quality is proactively managed as an ongoing process. Organizations have implemented advanced data quality tools and technologies, such as data profiling, data cleansing, and data monitoring tools. Data quality initiatives are integrated into the overall data governance framework, and there is a focus on prevention and continuous improvement.
  5. Optimized Stage: The highest level of data quality maturity is the optimized stage. Organizations at this level have a well-established data quality program with mature processes, tools, and technologies. Data quality is ingrained in the organizational culture, and there is a strong emphasis on data stewardship and accountability. Continuous monitoring, measurement, and optimization are routine and integral to the organization's operations.

of data quality are key aspects of the data management strategy.

It's important to note that the specific levels and their descriptions may vary across different data quality maturity models. Organizations can choose or customize a model that best suits their needs and goals. The primary objective is to progress from lower levels of maturity to higher levels by implementing appropriate data quality practices and capabilities.

Regenerate response

- +++++
- We will be reviewing the progress with Liv on Friday. So we need the following – **Dave - review of draft end of day 06/06**
    - One summary slide – 4 streams of work (not at the task level, more summarized milestones) with two columns 1) expected outcome during analysis ,2) expected outcome post final delivery – ( This should directionally align with SOW /deliverables )
    - One page for each of the effort – 4 quadrant slide covering ,
      - Summary of analysis and discovery,
      - Discussion and meetings held ( details with whom and bullet or two of synopsis)
      - Actions to be covered during the detailed execution phase
      - Any open item, follow ups, issues risks to be closed during execution
    - Overview of work for next 3 week,
      - High level tasks – dates
      - next review – (proposed agenda)

Summary slide - 4 streams of work at

4 streams of work (at summarized milestones level) with two columns 1) expected outcome during analysis , 2) expected outcome post final delivery – ( This should directionally align with SOW /deliverables )

Work Stream	Expected outcome during analysis	Expected outcome post final delivery
Use case analysis	Selected proactive use cases	DQ implementation plan for TWO focused data quality issues with go-forward processes for proactive issue identification and remediation
DQ Tool analysis	Identified a potential DQ tool for the Global ID replacement	DQ Tool Recommendations to replace Global ID
SDLC Process analysis	Analyzed current Process Intake template(s) and made appropriate changes recommendations Have identified KPIs to monitor and measure to easily assess the DQ.	Documented SOPs and implementation plan to incorporate DQ into software development and EA process KPI's to monitor and measure the business and technology adoption of DQ for ONE inflight project
Metadata Lineage	Data Flows and data attributes have been identified for the selected use cases: Smoking Surcharge, Plan add, Premium Payments. Attribute mapping to data sources has started	Metadata lineage for THREE selected use cases for AOP / COMPAS migration. Do we have to implement metadata lineage tool??

#### Process Intake process: [Portfolio Hub - Portfolio Intake Methodology \(sharepoint.com\)](#)

**Strategic Idea Expansion**  
The proposed idea is further defined, including specific information related to strategic alignment/value, sizing, and solution recommendation

[Expansion thru Feasibility Resource Assessment](#)  
[Presentation Template 01.18.23.pptx](#)

[Project Documentation](#)  
[Idea Expansion Spreadsheet Template\\_07052022.xlsx](#)

**Expansion Steps**

- Core Idea Team is assigned; in Aha! Core Idea Team includes Business Owner, Finance, Enterprise (Solution) Architect, Business Architect, Program Analyst, Program Manager, and SME's as needed
- Core Team completes Idea Expansion fields, with attached high-level solution diagram, cost-benefit analysis (CBA), and MVP scope details
- Project update: Assigned Program Manager provides status updates in Aha! Assigned Program Analyst expands on project information in Aha! based on Core Team input
- Finance department creates a CBA
- Core Team reviews completed Idea Expansion proposal with Strategic Choice Owner, Director of Portfolio Management and the Sr. Director of Business Transformation for approval
- Executive Champions team receives Idea Expansion for approval and initial prioritization for Feasibility and Resource Assessment (FRA)
- With Executive Champion approval for FRA, idea is promoted to initiative in Aha!

**Ideation: Expansion**

Core Team is assigned to the project

Core Team and Strategic Choice Owner review proposal

Strategic Choice Owner & Executive Champion Review

Capture relevant Scope, MVP, Value, T-Shirt Estimates, etc.

SCO, Business Owner & Program Manager review outcome with Portfolio Planning leadership

Executive Champions collectively review Idea Expansion proposal for approval and prioritization

Next Stage

Parking Lot

Reject Idea

Status Updates and Project Details incorporated into Aha!

**Evaluation: Feasibility & Resource Assessments (FRA), Final Prioritization and Roadmapping**  
Idea is evaluated for organizational impacts and resource assessment. The idea is then roadmapped and scheduled for execution/kickoff

[Project Documentation](#)



## EZ Claim Pay Phase 2

### Ideation Review/Approval Phase



October 28, 2022

United  
Healthcare

## Strategic Framework

**Vision**  
Be a trusted partner for older Americans on their health and wellness journey



NORTH STAR  
VALUE DISCIPLINE

### Mission

Deliver:  
• Supplemental Medicare coverage you can count on,  
• Simple experiences, and  
• Extraordinary lifetime value

CONSUMER FOCUS  
65 NPS  
30% Market Share  
PROGRAM GOALS



### STRATEGIC CHOICE(S)

Expense Management Ben Schneeman	Admin Expense Ratio
Book Stability Steve Glassman	New Business Loss Ratio
Rate Competitiveness Xiaoping Hu	Rate Competitiveness Score
Product Portfolio Patrick Zampogna	Features / Value Score (MSQS)
Health and Wellness Annette Fellows	Plan Cares About Me (MSQS)
Consumer Demand Generation Reggie Menon	Marketing & Sales Effectiveness & Marketing Efficiency
DTC Shopping Joe Fitzpatrick	Direct-to-Consumer Sales
External Sales Partnerships Martina Wiedmayer	Agent & Employer Sales
Enrollment Monica Marchetta	Overall & Online Enrollment Scores (CSAT)
Onboarding Jocelyn Barton	New Member NPS
Service Dan MacLauchlan	Easy to Do Business With Score (CSAT)



<https://confluence.optum.com/display/DMSKB/Talend+FAQ>

New idea | Insurance Solutions Portfolio Ideas Portal (aha.io)

### MRIS DQ Enablement

Problem statement

- Poor data quality and they lack a defined DQ process, and it's on-demand
- Very limited understanding and capability of metadata lineage
- Manual Metadata Lineage (track data movement, helps to understand data source and target, how it is used and transformed)

They engaged OAS to help them operationalize data quality and help design and define metadata lineage process  
Since there are two objectives and we have two parallel work streams -1. DQ operationalization; 2. Define the reusable metadata lineage template

Our approach

DQ

1. Use case analysis
  - i. Identify use cases to implement DQ operationalization
  - ii. Classify use cases Proactive vs. Reactive --> move data quality enablement as far left as practical
  - iii. Identify the process owners
2. Data Quality tool
  - i. Current DQ tool assessment is less than optimal, and it's been decided to be replaced
  - ii. Researched DQ tools that meet the DQ defined criteria
  - iii. Selected the top 3 DQ tools - they were DQaaS, Talend and OpenMetadata
  - iv. Arranged for vendor product demos
  - v. Created a template to capture user's DQ feature ratings (1-5)
  - vi. Distributed the template to the attendees and requested them to score each feature of the DQ tool as best as possible
  - vii. Receive user provided ratings to determine the best tool at the best price point

- viii. Recommend the DQ that will best meet MRIS' needs
3. SDLC process analysis
    - i. Reason: to understand the current Process Intake process and explore opportunities and be able to move the DQ as far left as pragmatic
    - ii. Met with EA, intake, Project governance and execution teams to understand where DQ operationalization could potentially be injected
    - iii. We learned 3 possible injection points -
      - 1) Architecture - Intake / solution guidelines
      - 2) **Project governance** – Project completion criteria
      - 3) **Project Execution** – Sprint / PI planning
    - iv. Need to define DQ metrics
    - v. We are creating reusable use case templates
    - vi. We are working with the EA personnel to implement DQ operationalization
    - vii. Pilot DQ enablement to implement SOPs
    - viii. Establish cadence for DQ monitoring
  4. For Metadata lineage
    - i. Business Metadata Lineage (BML) definition is uncharted area for the client
    - ii. **Attribute / Reporting object aligned business lineage can be of better business value**
    - iii. Business Metadata Lineage can leverage current technical metadata lineage being established
    - iv. **Playbook / template for Business metadata lineage**
    - v. Data Lineage template for reusability

Solution Intent document:

RE: Data quality questions

Barker, Scott D  
 To: McElroy, Jennifer; Ferguson, Kimberly  
 Cc: Cheema, Dave; Ramakrishnan, Anantha  
 Retention Policy: UHGIinbox (90 days)  
 Internal

Thu 6/15/2023 10:30 AM

Expires: 9/13/2023

Thanks for your time today Jenn. These turned out to be good set of examples.

Data Quality examples discussed:

1. Fox Plan Table Maintenance – there are many families of plan tables in Fox Claims – the Plan Type (PTIF\*), Plan State (PSIF\*), Cross Reference (product/plan, internal/external display, others). They can get out of synch. No official owner from business side. (people/process issue)
2. Partition Number – this is legacy column from UCPS, not used in Fox and should be set to 0 100% of the time. Non-zero instances will cause joins to fail. Mystery problem of non-0 instances just popped up.
3. Date of Death – two columns called DSCD\_DT in Fox Claims – ACCT\_INFO table and MBR\_ALT\_ADDR. Can get out of synch and some programs still use MBR\_ALT\_ADDR. Tech debt issue that it was not removed from MBR\_ALT\_ADDR. (known tech debt issue)
4. Work Queue/BPM tables – BPM is old custom built product from Red Hat many years ago. WQM tables are new work queue mgt tables in Fox. Hard to manage and coordinate these tables in Fox. No overall owner/coordinator for their maintenance. (legacy software and people/process issue)
5. BPM tables – we have table listings, and data model, but no definitions for BPM tables or columns. (metadata issue)

Please let me know if above is correct and please let me know if you think of any more. I'll send separate note on what I found for BPM documentation.

Scott

Notes 06/15/2023

o EA, We'll go in with our recommendations and then make adjustments

Ideation --> Idea Expansion - EA (Architecture Solution Intent document)

Solution Intent - SMS Text eKIT --> EA Solution Intent Repository

Governance --> Planning --> Cadence based Planning & Prioritization

Execution - Part of the EPIC refinement(utilize PPR output assessment and schedule kickoff)

Data accuracy, Data completeness, Data consistency, Data timeliness, Data accessibility, and Data relevance

EA - Put a slide for DQ in the David Searfass's de

Governan

## Execution

## Onboarding Kit checklist

Sujaya is the person to understand SQL and how problem data extraction

### **Table 1B: Quality**

Talend Data Quality vendor meeting 06/15/2023  
Subj: [REDACTED]

Only in commercial version

In it a service, or an executable that runs

Different licenses for on-prem and cloud  
https://www.microsoft.com/licensing/licenses/

Licenses are based on number of

## Things to complete

Write something

- Each link take you to some guideline - do some things at what spot.

  - I have created a new version ([Solution Intent - SMS Text eKIT v01 - DC.pptx](#)) of the deck that David Searfass sent.
  - In that deck, I have modified slides # 2, 5, 18 and created new slides #8, 11, 12, 13
  - I have also created:
    - [MRIS DQ Enablement -NFR-v01.1.docx](#) document that details the Data Quality NFRs.
    - [Data Quality Enablement Guidelines.docx](#) document
  - I have added a line # 39 (see below) to MRIS Enterprise Architecture Guidelines.xls:
  - |      |    |                    |     |   |
|------|----|--------------------|-----|---|
| MUST | IS | NFR - Data Quality | NFR | <a href="#">MRIS DQ Enterprise Architecture Guidelines.xlsx</a> |
|------|----|--------------------|-----|---|
  - Please note that Grouping is NFR

Please review them closely and make any and all necessary changes. Thank you.

## Use Case classification

## Use Case classification

Items after Date of Death

- Q. Where is member's information (Date of Death) come from and how often?
- Q. How does the Date Of Death gets out of sync in DSCD\_DT in Fox claims and in MBR\_ALT\_ADDR?
- Q. How often does the DOD notification is delayed?
- A. Sounds like a technical debt issue. Need to understand does the DoD in MBR\_ALT\_ADDR get cascaded? Is there a recon process? Is it event triggered, scheduled or something else?
- R. Sounds like a reactive use case. Before a claim is processed, if the DOS is later/greater than the DOD, reject that claim.

#### Simple use cases

Profiling agent contact data elements for basic DQ dimensions of completeness, format, consistency, etc. For example:

1. Phone numbers – how many formats; included dashes or parentheses; are all numeric, are 10 digits, missing phone, etc.

A.	MDM	INVALID PHONE NUMBER DATA	Invalid Phone Format	Tech Problem	phone numbers containing invalid data or invalid phone numbers shared by multiple members •Alphabets in Phone Eg.Phone "2013373300E" Ext "T334" Phone 8016738085(CELL) •Text in Extension Eg. "CELL", "SON" •Shared Phone Numbers 2151231234 0000000000 8125352971 9999999999 1111111111
----	-----	---------------------------	----------------------	--------------	---

- B. Proactive use case. Can be applied to the incoming dataset, before it lands in the at-rest place
2. State codes – Is state code in the State code table; do state codes and zip codes align; missing state codes
  - A. Proactive use case. Can be applied to the incoming dataset, before it lands in the at-rest place. USPS [ZIP Code™ Lookup](#) dataset contains state codes and zipcodes

#### Complex use case:

1. Email (multiple people sharing email address, multiple members have same name and DOB - how do we distinguish?
  - A. Reactive use case. Should be part of a recon process.
2. Multiple address locations - how do we identify the most relevant

A.	permanent versus temporary addresses are not distinguished by SMART	Analyze each system for temporary and permanent residences based on business rules	Tech Problem	Issue reported by Marketing Leadership that permanent versus temporary addresses are not distinguished by SMART and this affects Retention Marketing and RTOs. • This is for any VAS (service) communication to members driven by SMART data where services differ between states. If the services differ in the different states, the wrong information can potentially be delivered. • The impact is being assessed to determine recovery actions in order to ensure the correct information is reaching its targeted audience.
----	---	--	--------------	---

- B. Reactive use case. Define rules what constitute a permanent vs. temporary address

## Tasks to do

### Use cases

- identify / List business impact of DQ - Scott to provide anecdotally
- Tabulate data elements, process(es) impacting data elements
- Review, refine with Scott B. and Kim F.
- Build the charter with KPI to track and implement

### Tool analysis & replacement

- Get demo from Talend Data Quality - scheduled
- Get second demo from Informatica - requested
- Score DQ tools
- Review with Dinesh and Scott separately
- Make DQ tool recommendation
- Guesstimate rollout timeline
- Document training option for the DQ tool
- Schedule time Gaurav, document steps for POC and with timelines and commitments
- Create block diagram of how DQ tool is working in two formats - on-prem and in the cloud

### SDLC

- Complete and review documents with EA
- Review with EA
- Complete Governance KPIs
- Review with Governance team
- Identification pilot project for DQ operationalization
- DQ enablement in SDLC and EA - in-progress
- Document the DQ enablement process (in SDLC and EA)
- Roadmap/plan for other ongoing projects
- Implement KPI cadence

- \* Review with Liv of June 30th or July 5th

### Talend:

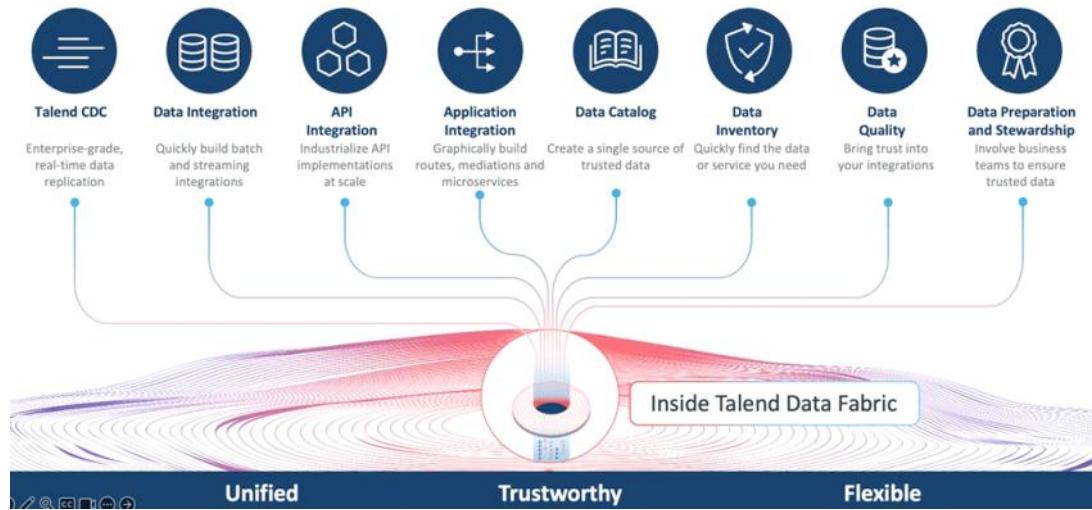
# Today's Agenda

- Shared Criteria Review
- Talend DQ Portfolio
- Demo #1: The DQ Launching Point
- Demo #2: The DQ Analyst Experience
- Demo #3: Operationalization of DQ Practices
- Talend Data Quality Framework (optional)
- Concluding Thoughts

Item#	Feature	Talend Strategy and Capability
1	Data Profiling and Assessment	Talend Data Preparation
2	Data Cleansing and Standardization	Talend Data Preparation
3	Data Types	Fully Customizable Semantic Types
4	Data Sources	Massive Quantity of Sources (1000+)
5	Data Quality Rules and Monitoring	Fully Customizable Rules Reported in Inventory
6	Machine Learning and Automation	Semantic Typing, Data Standardization, Matching, +
7	Data Integration and Connectivity	Studio / Pipeline Designer
8	Data Governance and Metadata Management	Talend Data Inventory (Upgradable to TDC)
9	Ease of Use and User Interface	Ground Up Designed Standard UI
10	Data Quality Dashboards and Reporting	Data Inventory (Upgradable to DQF)
11	Scalability and Performance	Ready for any enterprise topology
12	Cost and ROI	Constrained persona-driven pricing model
13	Licensing on-prem vs. cloud	Same on-prem vs cloud licensing model
14	Transferability of license	

## Talend Data Fabric

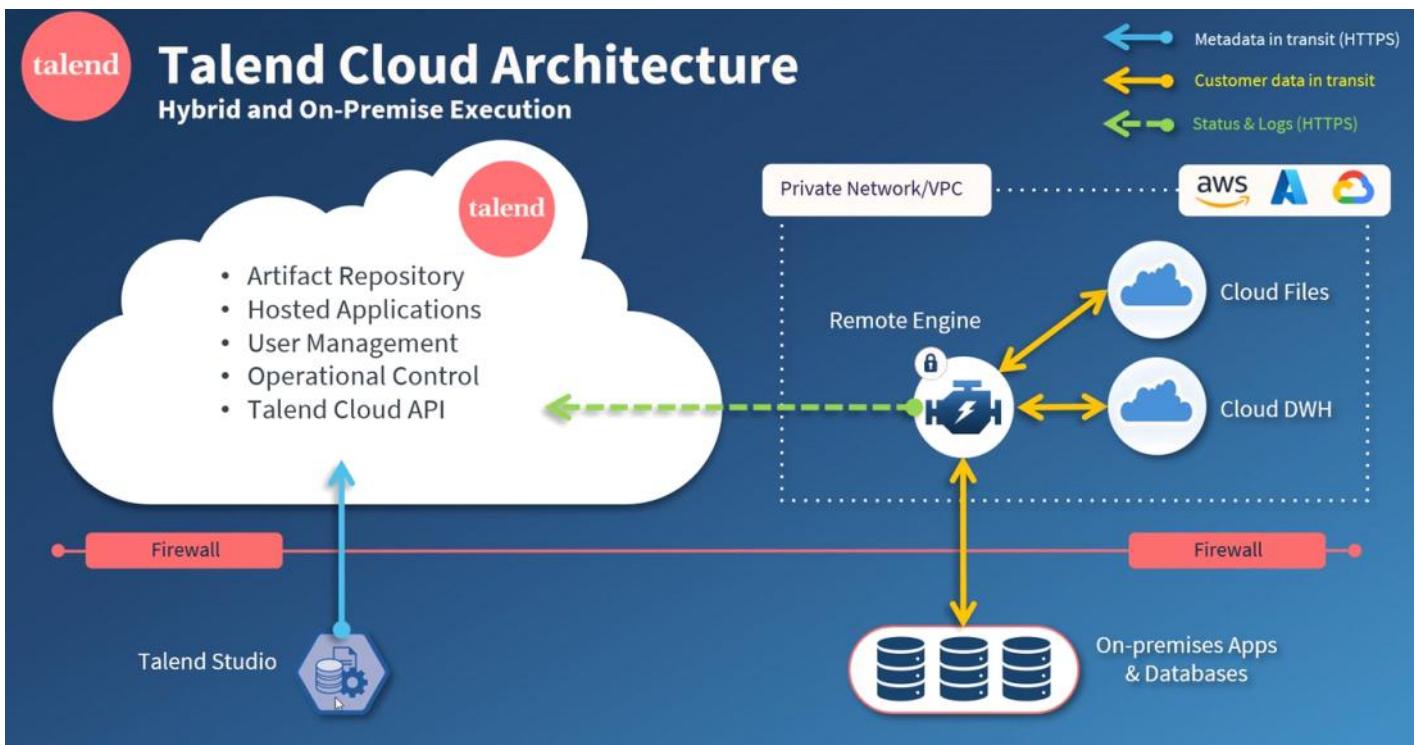
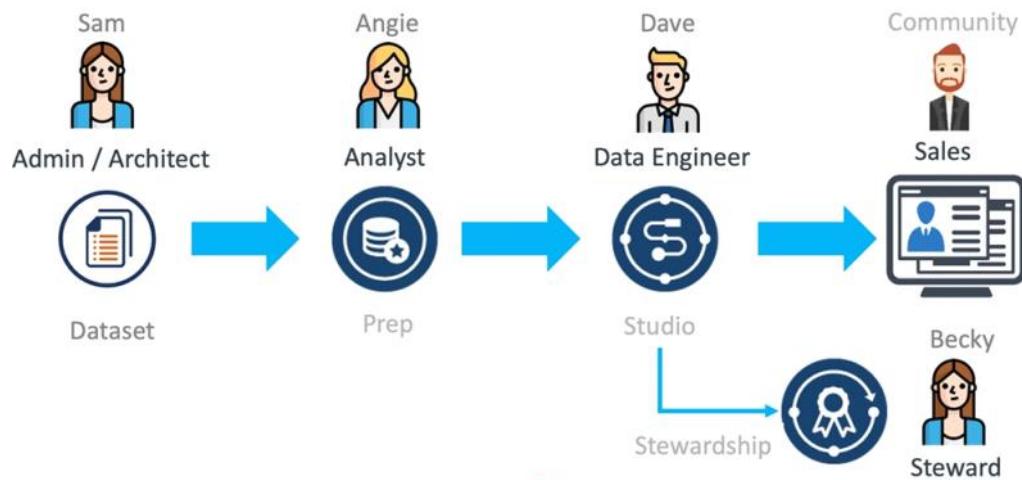
*Fully integrated, Always on, Self-Service, End-to-End Solution*

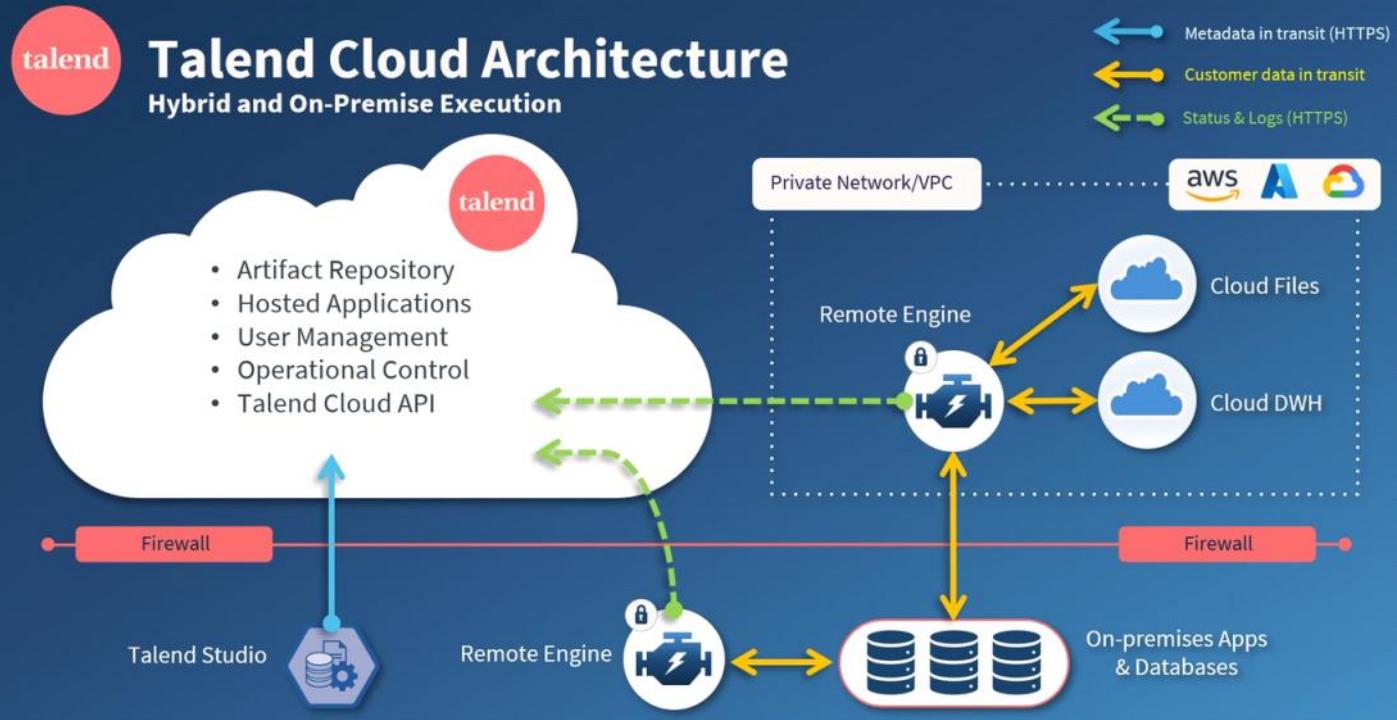


# The Talend Data Quality Product Strategy



## Data as a Team Sport





Notes 06/22/2023

Start review tasks with Scott to be done

Reach out Informatica again - how will be implement it

Informatica prep and ensure they have right team / next call schedule - done

Weekly status report - done

Tasks to do - done

Score the tools - done

KNIPEX 86 03 160

ICON 10 in. Pliers Wrench

Tasks to do

Use cases

- Identify / List business impact of DQ - Scott to provide anecdotally
- Tabulate data elements, process(es) impacting data elements

Tool analysis & replacement

- Get demo from Talend Data Quality - scheduled
- Get second demo from Informatica - requested
- Score DQ tools
- Make recommendation
- Guesstimate timeline for rollout
- Training option for the DQ tool, DQ enablement in SDLC and Architecture processes

SDLC

- Identification pilot project for DQ operationalization
- DQ enablement in SDLC and EA - in-progress
- Document the DQ enablement process (in SDLC and EA)
- Roadmap/plan for other ongoing projects

Rose wants DUR to be filled out

They need a sponsor e.g., Sojan Paul and Scott Barker

Scott

- Idea created:HCPDATAINT-I-422 -Setup DQaaS and a POC
- Review Data Quality section of the MRIS DQ Enablement -NFR-v0.1
- Review the Scoring of DQ tools comparison

06/21/2023

Met with DQaaS and created a work intake request to start planning the DQaaS service setup and create a POC

Sent Scott DQ tools comparison scoring, Use case analysis documents, MRIS DQ Enablement NFR, Data Quality Enablement Guidelines,

Resolved the need for DUR - no need

Optum RX

VP - Data Architecture blue print to present to her

Scope outcome in two box

data support

OptumRX Data Platform Strategy

Data Platform Strategy

Data strategy blue printing

[7:33 PM] Ramakrishnan, Anantha

[3:50 PM] Shin, John S

We have an opportunity with OptumRx to lead the definition of Data Platform Strategy in partnership with the Chief Architect of OptumRx. This effort will help us lead data strategy blue printing and will help with subsequent opportunities.

The staffing for this opportunity is for 2 senior resources for the remainder of 2023 – 6 months. Anantha will lead the initiative @60% as the strategic lead and Dave Cheema will be 100% dedicated as the data platform architect. Dinesh will be providing executive oversight @ 4hrs/mth to ensure there is alignment at the leadership level with OptumRx.

This engagement will be T&M with no contingency but the targeted financials for the engagement is shown below. Please let me know if you have any questions or anything to discuss. Otherwise, please see approval instructions below.

Rough timeline what we'll do and milestones

+++++

## DQ Operationalization

### DQ Tool

Once the DQ tool is selected and its buy-in is received, follow the steps listed below:

To work with the DQaaS team -

#### Tenant Onboarding

At a high level, the tenant on-boarding journey can be divided into the following steps:

1. Intake -
  - A. Initial step where a tenant interested in DQaaS raises an Aha request - done
  - B. Intake request gets attended by the Product owner, who engages with the tenant offering demos / information & capturing requirements.
  - C. Tenant can choose to submit their requirements in Aha against DQaaS capability to ensure there is a tracking on the customers asks and visibility into the roadmap.
2. Tenant console setup -
  - A. Request access to AD groups and setting up tenancy
  - B. Administrator in DQaaS will manage members of the tenant
3. Tenant Execution setup -
  - i. Covers all necessary aspects for the tenant to be able to create rules and run jobs.
  - ii. Requires some engineering effort from the tenant to be able to provision the required resources and configure the environment to run DQ jobs.
  - iii. DQaaS provides utilities & automations to minimize the manual effort; however, environment configurations cannot be fully automated.

#### Intake

- a. Pre-requisite -
  - i. Access to Aha Portal or register [here](#)
- b. i. Intake request -
  - i. Raise intake request in Aha to convey the need for DQaaS
- c. ii. Additional details -
  - i. After the demo and agreement to onboard, please attach the below excel (*request this Excel sheet from DQaaS coordinator*) to Aha intake.

#### Deployment model

Analyze and understand your use case and understand the available deployment options of the DQaaS and the appropriate model.

DQaaS is designed to operate in the following deployment models:

- A. **Near-data processing:** Move the processing close to the data instead of shipping the data to the processor. The DQ runtime runs on tenant's premise to connect to their storage systems.
- B. **Software-as-a-service(SaaS) model:** The common services are made centrally available as a SaaS model for the users to subscribe and access the DQ application.
- C. **Integration flexibility:** Tenants are provided with DQ runtime as an executable that can be integrated or called in data pipelines, from orchestration tools & other trigger mechanisms.

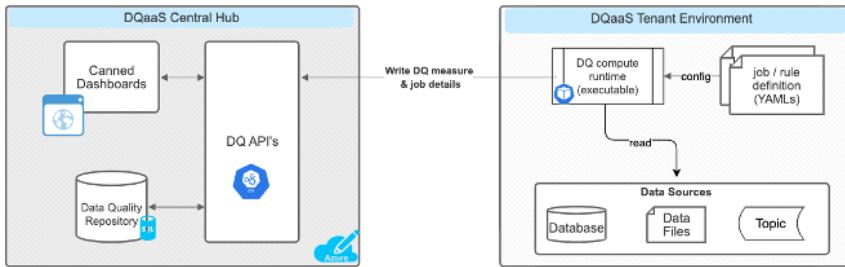
## SDLC

- Make DQ a project success criteria
- To enable DQ operationalization, there are two most logical entry points where the DQ enablement should be injected
  - in Idea Expansion (Solution Intent document)
  - At the beginning of the Project Governance (Planning --> Cadenced based planning & prioritization)
- Update / Create (**Governance template(s) to be identified**) templates
- Process owners verify the planning and execution of DQ Operationalization
- At this point, DQ enablement should be part the project SDLC, where it will go thru all remaining phases of the SDLC

## Use cases

- Approach should be that **DQ is required**.
- The approach should be to **move DQ enablement as far left as pragmatic**.
- Emphasis should be to **implement DQ proactively** and do it reactively where it wouldn't add any value to implement it proactively.
- Analyze the SDLC pipeline to determine where DQ enablement is most appropriate. **Note:** there can be **multiple entry points** in the SDLC pipeline where DQ could be injected.
- Based DQ requirements, configure DQ rules, thresholds and notifications
- Export configurations into a YAML file
- Setup an execution job and schedule it based on your use case
- Run the job, review results, and take appropriate measures to remedy the problems

# DQaaS Architecture



Key components & services of DQaaS:

- DQ compute runtime:** The DQ compute runtime is a scalable and fault tolerant DQ executable that processes input rules on the configured data sets and generate the results. The DQ runtime is built to run on any cloud provider Kubernetes service and recommended running in tenant's environment to move the compute closer to where the data is residing. The DQ runtime is callable and run DQ jobs in ways such as - integrate to their data pipelines, schedule via CronJob or job orchestration tools such as Airflow or Oozie.
- Declarative YAML files (job / rule):** The YAML declarative files carry the tenant configuration that feeds the DQ runtime to run jobs & rules & data source to run on. These files can be created manually or generated out of the UI.
- DQ API's:** DQaaS provides API endpoints for the tenants to programmatically access their DQ results. It provides mechanism to access and persist the information in central repository.
- DQ Central Repository:** A central database that stores the DQ rules, measures, aggregated scores & job details to facilitate quality analysis and reporting; does not store actual data or column values.
- Canned Dashboards:** DQaaS provides tailored dashboards to help tenants to monitor & analyze the quality of their data sets. They are available on the console where tenants can configure jobs, rules, checks & alerts.

## Deployment model

DQaaS is designed to operate in the following deployment models:

- Near-data processing:** Move the processing close to the data instead of shipping the data to the processor. The DQ runtime runs on tenant's premise to connect to their storage systems.
- Software-as-a-service(SaaS) model:** The common services are made centrally available as a SaaS model for the users to subscribe and access the DQ application.
- Integration flexibility:** Tenants are provided with DQ runtime as an executable that can be integrated or called in data pipelines, from orchestration tools & other trigger mechanisms.

## Tenant Onboarding

At a high level, the tenant on-boarding journey can be divided into the following steps:

1. Intake -
  - Initial step where a tenant interested in DQaaS raises an Aha request - done
  - Intake request gets attended by the Product owner, who engages with the tenant offering demos / information & capturing requirements.
  - Tenant can choose to submit their requirements in Aha against DQaaS capability to ensure there is a tracking on the customers asks and visibility into the roadmap.
2. Tenant console setup -
  - Request access to AD groups and setting up tenancy
  - Administrator in DQaaS will manage members of the tenant
3. Tenant Execution setup -
  - Covers all necessary aspects for the tenant to be able to create rules and run jobs.
  - Requires some engineering effort from the tenant to be able to provision the required resources and configure the environment to run DQ jobs.
  - DQaaS provides utilities & automations to minimize the manual effort; however, environment configurations cannot be fully automated.

## Intake

1. Pre-requisite -
  - Access to Aha Portal or register [here](#)
2. i. Intake request -
  - Raise intake request in Aha to convey the need for DQaaS
3. ii. Additional details -
  - After the demo and agreement to onboard, please attach the below excel to Aha intake.

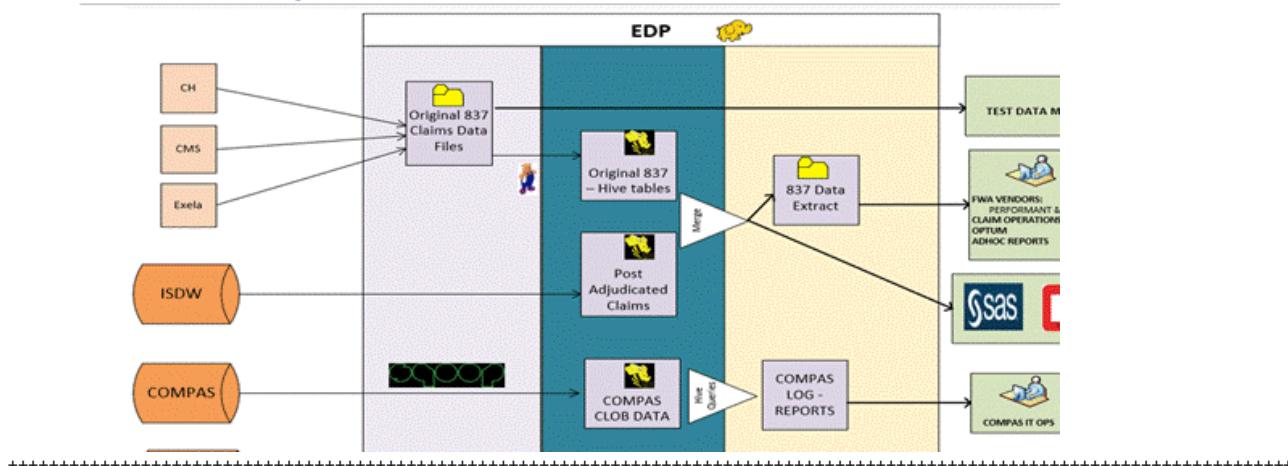
## Current state DQaaS Services & components

- DQ computation engine - performs metric computation when rules are applied and persisted in a database for reporting. It has few pre-built data source connectors, pre-defined checks (i.e. [14 checks](#)).
- UI - Intuitive interface that supports multi-tenancy model and allows users to - define rules, create & schedule jobs, visualize DQ metrics - at a high level.
- Data Profiling - This has a separate UI, API's & components

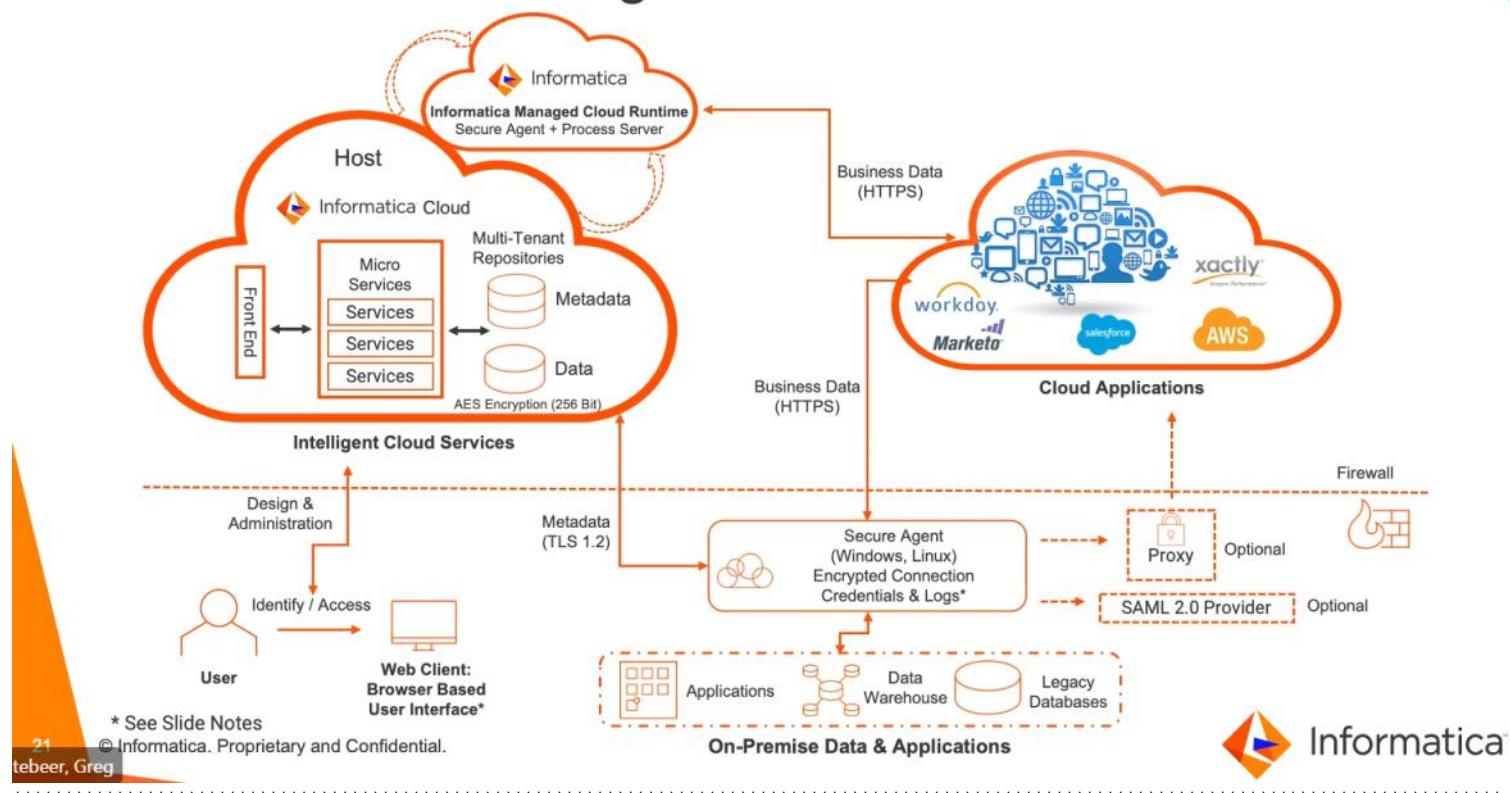
## Future state DQaaS roadmap

- Enhance codebase to offer an executable that acts on declarative inputs in federated implementations
- Reduce existing tech debt - modularization & optimization
- Improve extensibility for the teams to extend and expand
- Add checks above & beyond the 14 pre-defined checks
- Expand the data source connectors

**EDP is now comprised of several value offering solutions to our business partners**



## IDMC Architecture Diagram



UHG Reference Guidelines: <https://architecture.uhg.com/docs/enterprise-technology-strategies-detailed.html#data-management>

https://architecture.uhg.com/docs/enterprise-technology-strategies-detailed.html

COMPAS 828 LEET P Lucid BizzDesign ArchiSpecs UHG EA PDS UW Execution Port Intake Portfolio Execution Ideation SP GHEC IS AHA

UHG Enterprise Architecture Docs GitHub Enterprise

## Welcome to Enterprise Architecture >

### Application and Technology Architecture >

### Architecture Governance >

About EAG

Enterprise Technology Strategies

Enterprise Technology Strategies - Detailed

EA Workload Placement Guidance

### Federated Enterprise Architecture >

### Key Engagements >

### History >

## Enterprise Technology Strategies - Detailed

EA traditionally talks about "Architecture Principles" - those hallowed guardrails that assert the values of an organization. But they tend to be static, and constrain us towards a status quo. In contrast, strategies articulate direction, motion. Strategies drive action towards a goal or objective.

Our Enterprise Technology Strategies are Modernization, Leveraging our Scale, Data Management, Security, Digital Experience & Business Value, and Technology Innovation.

### Modernization

UHG must modernize its overall technology environment to adapt to changes in the business environment and allow for improved digital transformation, greater and faster business value delivery, and building of resilient / scalable solutions. This should be accomplished in a way that enables UHG to decouple solutions from its legacy application systems, reduce cost of healthcare, allow for continuous growth, and effectively compete with both established competitors and new entrants.

- **HealthCare Platform (HCP):** The Catalyst for change driving UHG to become a platform company with the aim to host all digital properties, business capabilities, and underlying data across the health spectrum. The intent is to make enterprise data accessible through data externalization techniques leveraging streams, reusable APIs, standardized canonical data models, and self-service access. The externalization of data will also help insulate legacy assets allowing for easily decoupling from core business processes and decommissioning.
- Package APIs should be the globally (i.e., across all United Health Group) authoritative source for given data or action. Package APIs serve the important function of bringing together disparate systems that operate on similar domain objects and abstract that away for our API consumers. In other words, our API consumers using a Package API see separate underlying systems as one. The Package API construct is critical to delivering a coherent world-class end user digital experience for those who do business with United Health Group. Package APIs have the following characteristics:
  - Recognized by package owner as the authoritative API for the domain
  - Discoverable and consumed via Healthcare Platform (HCP) by being on HCP API Management
  - Unambiguous, non-duplicate with clear domain boundaries across a sub package hierarchy
  - Addresses broader UHG (UnitedHealth Group) use cases
  - Consistent data across consumption channels (portal, EDI (Electronic Data Interchange), call

### Modernization

Reuse & Simplify

Data Management

Security

Consumer Experience

Consumer Experience

Business Value

Engineering Excellence

Enterprise Tech Strategy - Scoring Criteria for ARB

General Guidelines

What does rating 1-5 signify?

Modernization

Reuse & Simplify

Data Management

Security

Consumer Experience

Engineering Excellence

### Items to conclude:

We need to bring the following items to conclusion:

- Finalization of candidate projects to enable DQ enablement and monitoring at solution stage – EA to incorporate
- DQ enablement success criteria for Projects - discussion with project governance team to operationalize DQ and define the required success criteria as part of SDLC
- KPI and cadence establishment by Data governance team for DQ operationalization enablement

- Project selection
- Have to decide whether have another meeting with Tom Jarosz and Tracy DiMateo
- Cost estimates for OpenMetadata - give some estimates
- Socialize our findings of the DQ tool with EA team
- Michael Hastie, David Searfass

Some of our other things such as, creating roadmap are dependent on their completion. Thank you.

Tool - research, demos, comparison, TCO analysis, data migration

Meeting 07/05/2023 with Tom and Tracy

Subject: Status update

Satisfied with the overall direction and progress

Need directional pricing numbers for informatica and Talend, based on 12 people for on-prem

Mention DQaaS as a Value service based on input from Rob Maggs

Pre-screen - Lead Analytics Engineer - Rajesh Naidu

Pre-screen - Lead Analytics Engineer - Sravan Jakkidi

Get the use cases aligned - Dinesh

Aman sakti Dhiraj for tool pricing based on existing contracts

How to help org. to become data driven platform org.

How to move to data as a platform

What would be the effect of it.

How smaller org. can use data platform how would they benefit

Drive the create the data platform

Hadoop

Health care

Genesis

Scala - procedural and object oriented

Project selection

Have to decide whether have another meeting with Tom Jarosz and Tracy DiMateo

Cost estimates for OpenMetadata - give some estimates

Socialize our findings of the DQ tool with EA team

Michael Hastie, David Searfass

Items to conclude:

We need to bring the following items to conclusion:

- Finalization of candidate projects to enable DQ enablement and monitoring at solution stage – EA to incorporate

- DQ enablement success criteria for Projects - discussion with project governance team to operationalize DQ and define the required success criteria as part of SDLC
- KPI and cadence establishment by Data governance team for DQ operationalization enablement

Some of our other things such as, creating roadmap are dependent on their completion. Thank you.

---

#### Use cases

identify / List business impact of DQ - Scott to provide anecdotally - done

Claims, Member Info., Agent Info.

Tabulate data elements, process(es) impacting data elements - to be done by Kim F.

- Kim F. to provide - in-progress

Review, refine with Scott B. and Kim F.

- initial review - done

- Final review - as soon as Kim is done

Build the charter with KPI to track and implement

Data Profiling - identified fields

Key dimensions

- Data accuracy: The percentage of data that is accurate.
- Data completeness: The percentage of data that is complete and contains necessary information.
- Data consistency: The degree to which data is consistent and conforms to a set of rules and standards.
- Data timeliness: Degree to which the data is up-to-date and reflects the current state of affairs.
- Data relevance: Degree to which data is meaningful and relevant for the intended purpose.

#### Tool analysis & replacement

Get demo from Talend Data Quality - done

Get second demo from Informatica - done

Score DQ tools - done

Scott review - done

Review with Dinesh and Scott separately - to be scheduled with Dinesh

Make DQ tool recommendation - ready to make recommendation

Guesstimate rollout timeline

Based on the activities listed below, it will take 2-4 weeks

- Setup 1-2 weeks
- Configuration and validation - 1-2 weeks
- Choose data quality tool - done
- DQaaS and MRIS will -
  - Install data quality tool
  - Configure data quality tool
  - Run data quality checks to identify errors and inconsistencies in your data
  - Monitor data Quality to ensure DQ remains accurate and consistent
  - Get buy-in from stakeholders
  - Start small
  - Use a test dataset
  - Seek out DQaaS help to get you started

Document training option for the DQ tool

For training DQaaS will conduct trainings, initial demo and share training videos

Also, the following documentation is available:

- DQaaS Product Main Page: <https://hcp.uhg.com/products/dqaaS>
- Tenant Onboarding: [https://dqaaS-docs.optum.com/docs/GettingStarted/onboarding\\_steps](https://dqaaS-docs.optum.com/docs/GettingStarted/onboarding_steps)
- Setup DQ Jobs in console: [https://dqaaS-docs.optum.com/docs/UserGuide/setup\\_DQ\\_job](https://dqaaS-docs.optum.com/docs/UserGuide/setup_DQ_job)
- DQaaS Product Documentation: <https://dqaaS-docs.optum.com/docs/Product/KeyFeatures>
- DQaaS Architecture: <https://dqaaS-docs.optum.com/docs/Architecture/architecture>
- DQaaS Playbook: <https://dqaaS-docs.optum.com/docs/TechnicalArticles/playbook>

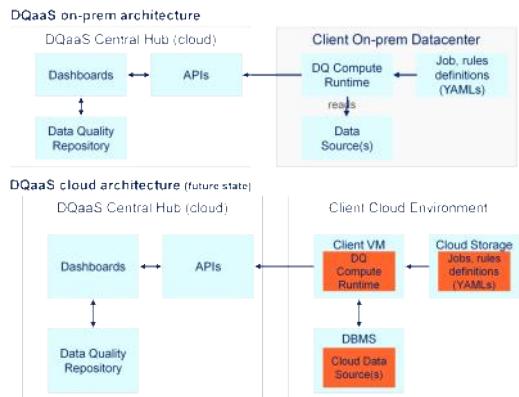
Schedule time Gaurav, document steps for POC and with timelines and commitments -

- Intake request - done

- work scheduled to start: July 15th, 2023

Create block diagram of how DQ tool is working in two formats - on-prem and in the cloud

#### DQaaS Architecture



Key components & services of DQaaS:

1. **DQ compute runtime:** The DQ compute runtime is a scalable and fault tolerant DQ executable that processes input rules on the configured data sets and generate the results. The DQ runtime is built to run on Kubernetes service and recommended running in tenant's environment to move the compute closer to where the data is residing. It is callable and run DQ jobs in ways such as - integrate to their data pipelines, schedule via CronJob or job orchestration tools such as Airflow or Oozie.

2. **Declarative YAML files (job / rule):** The YAML, declarative files, carry the tenant configuration that feeds the DQ runtime to run jobs & rules & data source to run on. These files can be created manually or generated out of the UI.
3. **DQ API's:** DQaaS provides API endpoints to programmatically access their DQ results. It provides mechanism to access information from and persist in central repository.
4. **DQ Central Repository:** A central database that stores the DQ rules, measures, aggregated scores & job details to facilitate quality analysis and reporting; but does not store actual data or column values.
5. **Dashboards:** DQaaS provides tailored dashboards to monitor & analyze the quality of their data sets. They are available on the console where tenants can configure jobs, rules, checks & alerts.

#### Deployment model

DQaaS is designed to operate in the following deployment models:

1. **Near-data processing:** Move processor (DQ compute runtime) close to the data, instead of shipping the data to the processor. The DQ runtime runs on tenant's premise to connect to their storage systems.
2. **Software-as-a-service(SaaS) model:** The common services are made centrally available as a SaaS model for the users to subscribe and access the DQ application.
3. **Integration flexibility:** Tenants are provided with DQ runtime as an executable that can be integrated or called in data pipelines, from orchestration tools & other trigger mechanisms.

#### SDLC

Complete and review documents with EA - scheduled for 06/28/2023

Review with EA - scheduled for 06/28/2023

Complete Governance KPIs - to be discussed with Anantha

Review with Governance team - scheduled for 06/28/2023

Identification pilot project for DQ operationalization - to discuss with EA and Governance teams

DQ enablement in SDLC and EA - in-progress

Document the DQ enablement process (in SDLC and EA)

Roadmap/plan for other ongoing projects

Implement KPI cadence



## Solution Intent Considerations

- Architectures and Solutions must be informed by UHG EA Standards, Insurance Solutions EA Standards, and Information Security Policy:
  - [UHG EA](#)
  - [Insurance Solutions EA Standards Library](#)
  - [Insurance Solutions EA Architecture Guidelines](#) - Updated / added the details in the Excel for DQ-
  - [HCP Engineering Handbook](#)
  - [Enterprise Information Security Policies](#)
  - [Data Quality Enablement Guidelines](#) - Added as new link / contents are subjected to refinement

## Business Objective

Enable real time SMS texting capabilities for the delivery of the eKIT link and the security code for Milliman release of information during online enrollment .....enabling proactive verification of data quality and information

## IT Impacts

High Level Business Ask	Anticipated IT Impacts	IT Estimate
Implement/configure the necessary short code(s) required for outbound SMS.	<b>Salesforce</b> - Configure Marketing Cloud Mobile Connect to use appropriate SMS short code	<b>Salesforce</b> - (X) 40 hours
Enable the Short Message Service (SMS) delivery channel for eKIT link and identity validation code for Milliman release of information.	<b>ACES/X</b> - Modify existing email service to deliver via the SMS channel - Modify data model to accommodate storing the SMS consent/preference - Modify data model to accommodate storing the SMS event - Build logic to handle negative scenarios (SMS delays, incorrect code, Salesforce outage, etc) - Build logic to capture SMS opt-out	<b>ACES/X</b> - (M) 480 hours
Persist the preference (or consent?) at the individual level. Must include timestamp (per LCR)		
Validate the security code when supplied on the submitted application(BAU).		
URL must be shortened to fit within SMS character limits.		
Change from six digit alphanumeric to six digit numeric.		
Capture applicant opt out, if applicable.		
Capture and persist communication event.		
Display option on QLE application for applicant to select SMS or email when receiving eKIT and security code.	<b>ACQ</b> - Invoke ACES email/sms service to send code and eKIT link to applicant	<b>ACQ</b> - (S) 240 hours
Capture outbound contact for either email or SMS and add to consumer 360 profile	<b>SMART</b> - Build new data model to store contact history event - Build new ETL to capture event	<b>SMART</b> - (M) 480 hours
(DQ Enablement ) - capture data that is complete, accurate and consistent. In addition, monitor, measure and report on data quality outcomes	<b>SMART</b> - Build new checkpoints in ETL / SMART for leveraging the DQ checks	<b>SMART</b> - (X) 60 hours

© 2022 United HealthCare Services, Inc. All rights reserved.

## DQ Enablement Considerations / Assumptions

### Data Quality Considerations

Directional construct,  
subject to refinement.

- Data quality measurement – are standards, DQ metrics and monitoring established?
- Data consistency – Is there a process to identify and correct data ?
- Data usability – Is it easy to extract insights from the data captured and is complete?

### Data Quality Assumptions

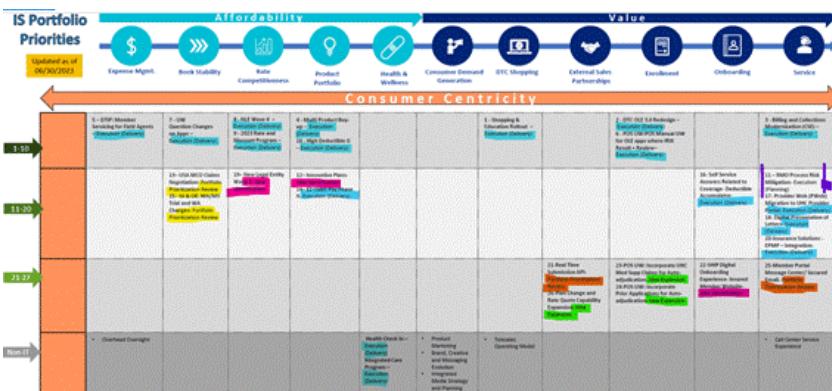
- Data is accurate and complete
- Data is consistent and up-to-date
- Data is meaningful for decision-making

## Non-functional Requirements

Requirement	Description	Approach
<b>Multi-tenancy</b>	Describe the requirement	Describe how multiple tenants or clients will be segregated.
<b>Upgrade or Update</b>	Describe the requirement	Describe how the architecture supports your upgrade or update requirements.
<b>Availability</b>	Describe the requirement	How will the architecture support the availability requirements?
<b>Data Quality</b>	Describe Data Quality KPI requirements	How will the architecture / DQ support monitoring of the DQ enablement
<b>Capacity</b>	Describe the requirement	How will the architecture support the Capacity requirements?
<b>Continuity</b>	Describe the requirement	How will the architecture support the Continuity requirements?
<b>Regulatory</b>	Describe the requirement	How will the regulatory requirements be met by the architecture?
<b>Provisioning</b>	Describe the requirement	How will the users and/or the application be provisioned?
<b>Automation</b>	Describe the requirement	Describe how the architecture maximizes automation and CI/CD

 © 2022 United HealthCare Services, Inc. All rights reserved.

+++++



Priority	Idea Name	Status	DQ Engagement*	Comments
11	RMO Process Risk Mitigation	Execution (Planning)		
12	Innovative Plans	Idea Identification		
13	USA NCO Claims Negotiation	Portfolio Prioritization Review		
15	GI & OE: MA/MS Trial and WA Changes	Portfolio Prioritization Review		
19	New Legal Entity Wave 5	Idea Identification		
21	Real Time Submission API	Portfolio Prioritization Review		
22	SHIP Digital Onboarding Experience-Insured Member Website	Idea Identification		
23	POS UW: Incorporate UHC Med Supp Claims for Auto Adjudication	Idea Expansion		
24	POS UW: Incorporate Prior Application For Auto adjudication	Idea Expansion		
25	Member Portal Message Center/ Secured Email	Portfolio Prioritization Review		
26	Plan Change and Rate Quote Capability Expansion	Idea Expansion		

\*Acceptable values: Yes, Not Qualified

Scott

- What were the steps/thoughts that lead you to believe whether DQ Enablement is not required? Please explain
  - Are you facing data quality challenges such as Inaccurate/incomplete reports, customer dissatisfaction or compliance issues? - yes
  - Does the content require to governed? - yes
  - Does the content create datapoints already being captured? No
    - If yes, is the end-to-end impact analyzed for down-stream systems? - No
  - Do you understand data quality needs for critical business data or any specific requirements for the data? - Yes
  - Are resources and processes in place to ensure data quality such as data steward, data quality plan, and tools to support data quality initiatives? - Yes
  - Are your personnel trained to contribute/improve data quality? Yes

+++++  
RACI - Responsible, Accountable, Consulted, and Informed.

SDLC Phase	Enterprise Architecture	Project Governance	Data Governance	Project Execution
Solution Intent / Idea Expansion	R, A	C	C	I
Portfolio Prioritization	I	R, A	C	I
Project Execution	I	C	C	R, A
Production Enablement	I	I	R, A	

- +++++  
  - Is data accurate, complete, and consistent?

- Is data protected from unauthorized access, use or disclosure?
- Is the methodology to identify PII and PHI established?
- Is metadata lineage established for PII and PHI for AARP contracts
- Does data comply with all applicable laws and regulations?
- Is accountability for the management and use of data clearly defined?
- For inbound data from external source(s) – have you created a B2B document?
- For outbound data, esp. PII data – Have you created a DUR document?
- Has retention policy for the data been defined?

\*\*\*\*\*  
put together a table of what is done, how is enabling DQ ops. Open items and who does it when

# Data Collection

Thursday, May 18, 2023 7:07 PM

## Deliverables

- DQ implementation plan for TWO focused data quality issues with go-forward processes for proactive issue identification and remediation
- DQ Tool Recommendations to replace Global ID
- Documented SoPs and implementation plan to incorporate DQ into software development and EA process
- KPI's to monitor and measure the business and technology adoption of DQ for ONE inflight project??
- Metadata lineage for THREE selected use cases for AOP / COMPAS migration. Do we have to implement metadata lineage tool

## DQ implementation plan for TWO focused data quality issues with go-forward processes for proactive issue identification and remediation

- QA enablement for every project is required. If not, then an exception justification is required
- Two potential areas in the SDLC have been identified where DQ enablement can be injected and ensure it is a project success criteria
  - EA (Ideation Expansion) and project governance (Planning --> Cadenced Based Planning & Prioritization) --> **To be validated**
- DQ in the EA process (Solution Intent document) have been identified
  - Awaiting EA (David S.) review and approval
  - Meeting with EA (David S.) has been scheduled for 06/28/2023
- DQ in the Project Governance process ([document/template?](#)) to be identified.
  - Meeting with David Weiss and team has been setup for 6/28/2023
  - Identified document(s)/template(s) will be updated accordingly
  - David Weiss' review and approval will be scheduled
- Candidate use cases have been identified
- Extract DQ requirements
- Identify DQ rules based on requirements
- Export DQ rules into a configuration (YAML) file
- Run DQ job
- Monitor and review results
- Take appropriate remediation steps

## DQ Tool Recommendations to replace Global ID

- Determined the need to replace the current DQ tool - does not incorporate data management into lifecycle, struggle to reconcile/implement Global ID
- Created a set of MRIS required features
- Researched and identified most appropriate DQ tools - DQaaS, Informatica, Talend, and OpenMetadata
- Compared tools' capabilities side-by-side and calculated weighted scores
- Determined DQaaS would be the most appropriate tool for MRIS DQ needs
- Recommend and receive MRIS buy-in
- Create DQaaS onboarding steps and procedures

## Documented SoPs and implementation plan to incorporate DQ into software development and EA process

To be created as soon as EA and Governance reviews and approve the DQ injection points in the SDLC

## KPI's to monitor and measure the business and technology adoption of DQ for ONE inflight project

**To be discussed with Anantha** - What I understood from was that this metric pertains the adoption of DQ in the SDLC.

- Q. How do we monitor and measure the adoption of this process?
- Q. How often do we do it and what mechanism do we use it?

## Metadata lineage for THREE selected use cases for AOP / COMPAS migration. Do we have to implement metadata lineage tool

[\*\*Portfolio Intake Methodology:\*\*](#) <https://uhgazure.sharepoint.com/sites/MRIS2/OCIO/initiatives/SitePages/Portfolio%20Intake%20Methodology.aspx>

### MRIS DQ stakeholders:

Paul, Sojan K (sojan\_paul@uhc.com)  
Scott Barker  
Rajkumar Thulasi  
Srilatha Paramatmuni  
Kim Ferguson, Shekhtman, Felix  
John Purcell (SDLC)  
David Weiss (SDLC, PMO)  
Jim Tannenbaum (EA)  
Mike Hastie (EA)  
Arjunan Easwaran

Use case Analysis	Status	MRIS DQ	Result	Comments
Identify all the current DQ use cases (agent, complex COMPAS email)	Complete	Scott B.		<p><b>Simple use cases:</b> Profiling agent contact data elements for basic DQ dimensions of completeness, format, consistency, etc. For example: 1. Phone numbers – how many formats; included dashes or parentheses; are all numeric, are 10 digits, missing phone, etc. 2. State codes – Is state code in the State code table; do state codes and zip codes align; missing state codes</p> <p><b>Complex use case:</b></p> <p>1. Email (multiple people sharing email address, multiple members have same name and DOB - how do we distinguish? 2. Multiple address locations - how do we identify the most relevant</p> <p><b>Note:</b> Agent is looking like a good use case for the DQ tool evaluation and process work. The volumes are manageable, no member PII, and data quality can vary.</p> <p>*****</p> <p>Date of Death use case:</p>

				<p><b>A</b></p> <p>1 Description (DoD Issue) 2 DoD data are not always being updated in COMPAS, which results in premium over payments, inaccurate financial reporting, and a negative member experience. 3 Process Improvement was implemented 2022 4 Obtain DoD information for active SHIP members from CMS and validating member data to be used for updating a member's account information. This new process helped reduce premium re-imbursements, produce more accurate financial reports and enhance member experience.</p> <p><b>DoD process flow</b></p> <pre> graph TD     A[Extract active SHIP members from the Plan Period] --&gt; B[Process the file]     B --&gt; C[Validate the file]     C --&gt; D[FTP file to CMS]     C --&gt; B     </pre> <p>16 Parameters for Use Case 17 Task: To build and analyze data for the below parameters to ensure the process is capturing all instances and included in the CMS transfer. 18 Report 1 19 Report with members who have a DoD from CMS and a claim record that occurred after their DoD (any dates of service that are after DoD/termination) 20 Dob, DoD, Surname, First Name, Membership Number, HICN, Gender. As well as a claim reference if possible, such as first date-of-service after DoD. 21 Report 2 22 Report with members whose DoD is after Paid through date 23 Separate records by employer associated records vs non employer accounts 24 Following data elements for Non-employer accounts: 25 Dob, DoD, Surname, First Name, Membership Number, HICN, Gender, Paid through date 26 Following data elements for Employer accounts: 27 Dob, DoD, Surname, First Name, Membership Number, Employer Name/Number, HICN, Gender, Paid through date, Joint Paid through date 28 Report 3 29 DoD &amp; Plan Term Date 30 C:\Users\kerau38\Desktop\DoD Flow.png</p>																						
Classify the reactive and proactive use cases	Not started	Scott B.		<p>Proactive use case: a business process designed <b>to prevent problem</b> from occurring. Reactive use case: a business process that occurs <b>after an event/problem</b> has occurred.</p> <p>Proactive use cases: Simple use cases (Phone numbers, State codes) Reactive use cases: Complex use cases (Email, Multiple addresses)</p> <p><b>SIMPLE USE CASE</b> Validation of Phone numbers, State codes process(es) should run before the data is loaded into the durable storage.</p> <p><b>COMPLEX USE CASE</b> Reactive – COMPAS email address - This process should run immediately after the COMPAS email address for over-merge process completes. Multiple address locations - This process should run immediately after the process that updates/adds the address location</p>																						
identify / List business impact of DQ	Not started	Scott B.		<p>Businesses impacted: Claims, Member Info., Agent Info., COMPAS, GPS, EIMP</p> <p>Business impact of DQ, such as, Actuarial, marketing, pricing, finance, enrollment, billing, etc. Contacting QA folks for info.</p> <p>Simple case – poor data quality of agent data would impact all reports in ISDW using it. Complex case – could cause member over-merge, resulting in member privacy breach.</p> <ul style="list-style-type: none"> <li>• Data Availability</li> <li>• Timeliness</li> <li>• Accessibility</li> <li>• Consistency</li> <li>• Comprehension</li> <li>• Accuracy</li> <li>• Completeness</li> <li>• Trustworthiness</li> <li>• Data Sharing</li> </ul>																						
Tabulate data elements, process that impact the elements	Not started	Scott B.		Data elements that are having quality issues and the processes that are causing those quality issues for those data elements.																						
identify data / Process owners (will ask around) o Sujaya Linga FOX (pre-batch data quality checks) Tina Tharakan Fox QA	Not Started	Scott B.		<table border="1"> <thead> <tr> <th>Process</th><th>Owner</th></tr> </thead> <tbody> <tr><td>MDM Quality Validation</td><td>Pam Hoffman</td></tr> <tr><td>Fox Pre-batch quality checks</td><td>Sujaya Linga</td></tr> <tr><td>SMART source data quality check</td><td>Verma</td></tr> <tr><td>EDP</td><td>Doug U., Sheharyar</td></tr> <tr><td>DEEP</td><td>Gurtej, Doug U.</td></tr> <tr><td>ISDW</td><td>Doug U.</td></tr> <tr><td>Fox Claims</td><td>Jenn McElroy, Brian Shafer</td></tr> <tr><td>Fox QA</td><td>Tina Tharakan</td></tr> <tr><td>COMPAS</td><td>Adam Cimino</td></tr> <tr><td>AOP project (in-flight)</td><td>??</td></tr> </tbody> </table>	Process	Owner	MDM Quality Validation	Pam Hoffman	Fox Pre-batch quality checks	Sujaya Linga	SMART source data quality check	Verma	EDP	Doug U., Sheharyar	DEEP	Gurtej, Doug U.	ISDW	Doug U.	Fox Claims	Jenn McElroy, Brian Shafer	Fox QA	Tina Tharakan	COMPAS	Adam Cimino	AOP project (in-flight)	??
Process	Owner																									
MDM Quality Validation	Pam Hoffman																									
Fox Pre-batch quality checks	Sujaya Linga																									
SMART source data quality check	Verma																									
EDP	Doug U., Sheharyar																									
DEEP	Gurtej, Doug U.																									
ISDW	Doug U.																									
Fox Claims	Jenn McElroy, Brian Shafer																									
Fox QA	Tina Tharakan																									
COMPAS	Adam Cimino																									
AOP project (in-flight)	??																									
<b>Data model analysis / Walkthrough</b>																										
<b>Tool Analysis</b>																										
Identify top 3 tools of choice for DQ	Not started	Scott B.		OAS to present option for MRIS DQ																						
Current TCO for DQ / Metadata tool	Complete	Scott B.		\$150,000/year + \$12K/month for the environments																						
Operating model, R & R	In-progress	Scott B., ??		What is the SOP today?																						
Projected demand	Not started	Scott B.																								
Data migration needs / volume / Set up estimation	Complete	Scott B.		No data migration needs, not volume to be migrated; just new tool set up estimation																						

Automation consideration	In-progress	Scott B.		Shared pre-batch DQ SQL statements
Shortlist data lineage business Use case (4 ) for AOP				
Tool for metadata finalization, Template finalization review				
SDLC Process				
Analysis of development process	Not started	Scott B., David Weiss		
Current gaps for DQ / DIG in development process	Not started	OAS		
Analyze Architecture process	Not started	Scott B., Jim T		
Gap identification for DQ / DIG in Architecture process	Not started	OAS		
Identify the beta projects for DQ operationalization (2 or 3)	Not started	Scott B., David Weiss??		
Review meetings				
Identification and documentation of all source / target data elements for identified use cases				
Sample lineage review				

MRIS DQ Enablement NFR requirements document: [MRIS DQ Enablement -NFR-v0.1.docx](#)

\*\*\*\*\*

#### Informatica Response:

Item #	Feature	Comments	Ratings (1-5)
1	Data Profiling and Assessment	Complete Profiling capability with drilldown from column to detail	5
2	Data Cleansing and Standardization	Create cleansing rules and apply directly to profile and re run to see improvements in iterations, then apply those reusable rules to your integration runs or analytics.	5
3	Data Types	Most anything structured or unstructured	5
4	Data Sources	As above, we can source from databases, applications, systems such as mainframes and midrange, unstructured data from Hadoop, S3, ADLS, GCP, OCI, etc	5
5	Data Quality Rules and Monitoring	Over 25 years of experience gives you over 200 rules out of the box and industry accelerators. Full monitoring capability in the platform and visible in Lineage	5
6	Machine Learning and Automation	Claire AI/ML is pervasive thru the platform and will deliver insights in each profile as well as guidance for automation.	5
7	Data Integration and Connectivity	Data Integration and connectivity are at the heart of the platform. We integrate data from cloud, Apps, middleware, Data Stores, Social and web apps, and Enterprise systems. On prem or cloud. Data Quality is integrated in line with Integration.	5
8	Data Governance and Metadata Management	Informatica Cloud Data Governance is a complete Governance solution and integrated into the platform. Data Catalog is the metadata management service for scanning all of your infrastructure and providing detailed information on your data assets as well as the lineage you are looking for. There is also complete business side governance with terms, policies, lineage, and glossaries integrated with the catalog. Users can find data via technical or business context.	5
9	Ease of Use and User Interface	Cloud Data Marketplace from Informatica is an easy to use, searchable service for end users to "shop" for data. All user interfaces for the various services have the same look and feel and are very intuitive.	5
10	Data Quality Dashboards and Reporting	Publishable Dashboards specific to Data Quality are available in multiple services. Reporting and Dashboards can be called in various ways via API	5
11	Scalability and Performance	Unlimited. You can leverage Elastic and Serverless infrastructure within your cloud ECO-Systems and Informatica will scale seamlessly. Processing is done "near the data" or with pushdown optimization within the databases. Claire can help determine the best path for a job.	5
12	Cost and ROI	With the complete platform at your fingertips, Informatica provides the most value by far in the industry. You have access to everything in our platform, all connectors and services and pay only by consumption of those services.	5
13	Licensing on-prem vs. cloud	All on-prem software has been "cloudified" at Informatica. You can still process on-premise as needed with our unique architecture.	5
14	Transferability of license	Yes	5

#### Talend Vendor evaluation

Item#	Feature	Talend Strategy and Capability
1	Data Profiling and Assessment	Talend Data Preparation
2	Data Cleansing and Standardization	Talend Data Preparation
3	Data Types	Fully Customizable Semantic Types
4	Data Sources	Massive Quantity of Sources (1000+)
5	Data Quality Rules and Monitoring	Fully Customizable Rules Reported in Inventory
6	Machine Learning and Automation	Semantic Typing, Data Standardization, Matching, +
7	Data Integration and Connectivity	Studio / Pipeline Designer
8	Data Governance and Metadata Management	Talend Data Inventory (Upgradable to TDC)
9	Ease of Use and User Interface	Ground Up Designed Standard UI
10	Data Quality Dashboards and Reporting	Data Inventory (Upgradable to DQF)
11	Scalability and Performance	Ready for any enterprise topology
12	Cost and ROI	Constrained persona-driven pricing model
13	Licensing on-prem vs. cloud	Same on-prem vs cloud licensing model
14	Transferability of license	

++++++  
Kim's Excel sheet of issues: [Data Quality Issues Log\\_2023.xlsx](#)

# DQaaS Overview

Monday, March 20, 2023 10:35 AM

DQaaS Product documentation: <https://dqaas-docs.optum.com/docs/Product/KeyFeatures>

DQaaS Architecture: <https://dqaas-docs.optum.com/docs/Architecture/architecture>

DQaaS Playbook: <https://dqaas-docs.optum.com/docs/TechnicalArticles/playboook>

DQaaS: <https://hcp.uhg.com/products/dqaas>

DQaaS Onboarding documents: <https://docs.hcp.uhg.com/dqaas/tenant-onboarding>

<https://optum.aha.io/ideas/ideas/HCPDATAINT-I-422>

The screenshot shows a web browser displaying the DQaaS Product documentation at <https://dqaas-docs.optum.com/docs/Product/KeyFeatures>. The page has a dark blue header with the "Healthcare Platform" logo, "Docs", and "GitHub" links. A navigation sidebar on the left lists categories like "Introduction", "GettingStarted", "Architecture", "TechnicalArticles", "Product" (with "Key Features" selected), "Roadmap", "DQaaS Customers", "UserGuide", and "Contact". The main content area features a large heading "Key Features" and a bulleted list of five features:

- **Content Data Quality Assessment:** DQaaS is designed to access the quality of the contents in the data across various dataset or products. DQaaS provides the list of diverse check for metric computation and constraint verification across various quality dimensions. DQaaS computes data quality metrics based on the checks and validations set and generates relevant reports
- **Quality metrics aligned to Key Data Quality dimensions:** Assessment of data quality is available across all key quality dimensions (**Completeness, Validity, Consistency, Uniqueness, and Integrity**)
- **Scalable self-service capability:** DQaaS provides a self-service capability which allow you to configure DQ checks or rule faster, more effective, and in a cost-effective way. This gives user ability to onboard to DQaaS without dependencies on anyone for onboarding to DQaaS.
- **Declarative way for developers' community:** This enables developer community to provide YAML based inputs to configure and run rules on their data assets. Now developers can create build rules and define checks at scale in an effective manner
- **Federated Approach model:** This approach helps customers to leverage their own compute environment and infrastructure to run DQ rules on their data sources. DQaaS runtime is capable is running in any Kubernetes environment either on cloud or on prem. This will remove any thread related to security concern due to data movement for computation.

https://dqaaS-docs.optum.com/docs/Product/keyfeatures

Rally AI in Market Resear... Aha Rally PPM optics TouchPoint: Meetin... HCDP Customer Int... Data Quality as a S... DQaaS-Dashboard HCP Console Docs Da

## Healthcare Platform Docs GitHub

- Introduction
- Overview
- Product Fact Sheet
- GettingStarted
- Architecture
- TechnicalArticles
- Product
  - Key Features**
  - Roadmap
  - DQaaS Customers
- UserGuide
- Contact

assessment due to federated and flexible approach.

- Detailed Dashboard for better visualization:** Data Quality score is available in DQ dashboard across various data quality dimensions. Other dashboards for checking the results score and failure for respective jobs and rules are available with drill down approach
- Data Profiling for extensive data analysis:** Data profiling is the parallel capability available for business/data analyst for detailed analysis of the data. It provides capabilities for analyzing, Null %, Blank %, Distinct count, min, max, most, and least frequently occurring values, minimum and maximum length, six categories of outliers (Numeric, length, Frequency, Data type, Pattern and Charter) on the contents of the dataset.
- Threshold based altering:** (Future) DQaaS got the ability to define multiple level of thresholds for user to define and generates alerts based the breach for the respective threshold. Alerting is provided to proactively send notification to data owners reflecting the failure scenarios. Customers can analysis for the DQ issues for various level of thresholds on DQ dashboard as well
- Rule Recommendation (Future):** DQaaS will suggest the recommended DQ checks to be applied on various data attributes leveraging the auto constraints suggestion and trained ML models
- Varied Data sources support with DQaaS** a. File based sources (Parquet, Avro, ORC, CSV, TSV, Text), b. JDBC based: Snowflake, MySQL, PostgreSQL, SQL server etc., c. Streaming platform: Micro batch (Non nested Kafka topics), d. Custom: Cassandra, e. Big Data platform: Hive

Meeting 04/05/2023 with DQaaS

Subject:DQaaS

Internal capability

Run locally in Kubernetes environment, can it run on a stand-alone server

Database access can be an issue, when you don't have proper access

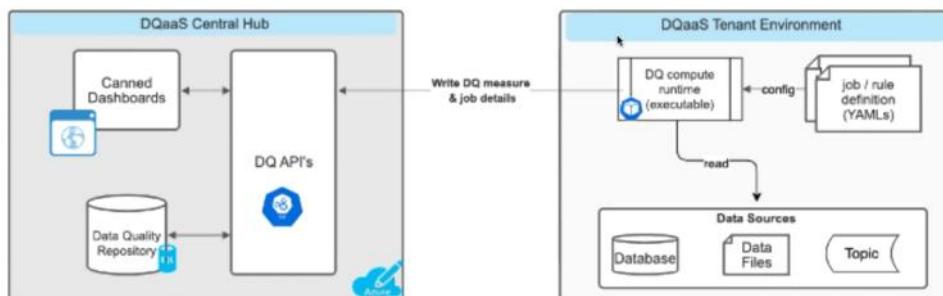
Who is responsible for the operations and maintenance

what is costing model - added to a group, per tenant - \$8771/month/tenant

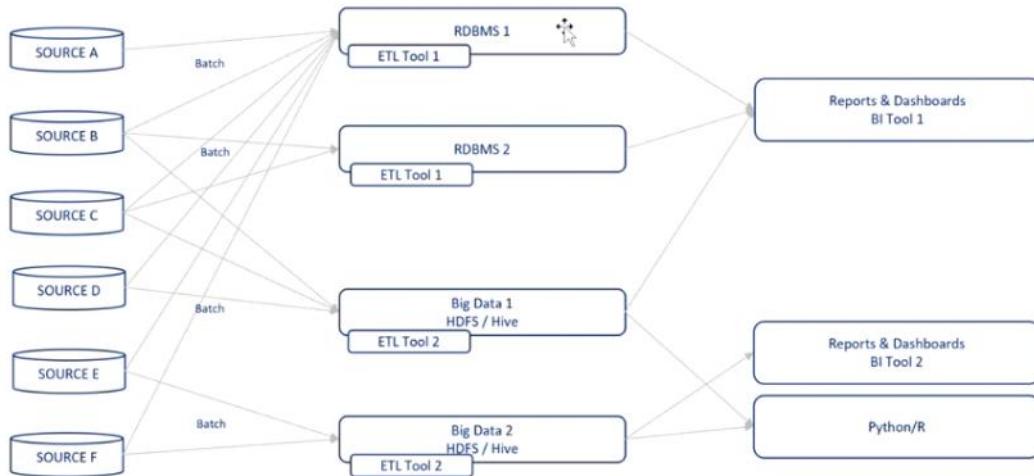
dqaaS-docs.optum.com/docs/Architecture/architecture

# Architecture Overview

This article covers the architectural details - components, services & deployment model - of DQaaS for tenants.



## **Current State**



→ C ⌂ 🔒 https://docs.hcp.uhg.com/dqaas/tenant-onboarding

(DQaaS) | Aha! Rally Team Turing Data Quality as a S... ServiceNow PPM Optics Chatgpt Self service : Global Talend review Intake Executive Dashboard

## Step 2 - Complete Tenant Registration

In this step, Tenant will complete registration & console setup to use the DQ services.

Following are the things to be taken care of:

1. To access DQaaS console, raise secure requests mentioning the 'Tenant name' for which you need the access to DQaaS, also mention the reasons and your Role with DQaaS for below AD groups:
  - Development & Stage - AZU\_DQaaS\_SSO\_Dev\_Users
  - Production - AZU\_DQaaS\_SSO\_Prod\_Users
2. Complete the API developer registration to interface with DQaaS API's & orchestrate your data pipelines. Follow instructions [here](#).

## Step 3 - Create Rules & Jobs

Follow detailed instructions located [here](#) to go through a step by step guide for creating rules & jobs.

## Step 4 - Export DQ Job as yaml configuration files

Follow detailed instructions located [here](#) to go through a step by step guide for setting up jenkins export pipeline.

## Step 4 - Deploy DQ Jobs

Key Features

Roadmap

DQaaS Customers

**Getting Started**

Tenant Onboarding

Capability Training Videos

Architecture

Technical Articles

User Guide

Product Support Services

Release Notes

Customer Testimonials

FAQs

# DQaaS Team Communications

Thursday, May 25, 2023 9:24 AM

## RE: Can you please compare DQaaS against these products



Agarwal, Gaurav [Optum]

To: Cheema, Dave

Retention Policy UHGIinbox (90 days)

Internal

You replied to this message on 5/24/2023 9:01 AM.

[Reply](#) [Reply All](#) [Forward](#) [Print](#) [...](#)

Wed 5/24/2023 6:44 AM

Expires 8/22/2023

Few others pointers in support

### Key Differentiator

1. **Easy to use interface**: Customized and simple interface for Business analyst and Quality Analyst to data quality governance ( Define rules, Jobs and view results with role based access control )

2. **Support Analyst and Developer Persona**: Apart UI , developers has the flexibility to write more complex rules ( Cross column and cross table level validations )

<https://dqaas-docs.optum.com/docs/TechnicalArticles/playbook>

3. **Native support for enterprise capabilities**

- a. Support most of the enterprise sources ( Hive, Files, RDBMS, Hive, Snowflake, Kafka ( non-nested ) , Kafka ( nested Avro ) : in progress, Databricks ( in progress), cloud bases sources
- b. Integration on data catalog : Q3-Q4 2023

4. **High performance benchmarking**: Saving of compute resources in longer run

	Executor Memory	Executor Cores	Connector Type	Processing time (mins)	Total Records
4	12GB	2	SNOWFLAKE	13.27	1.2 Billion
6	12GB	2	SNOWFLAKE	9.13	1.2 Billion
4	12GB	2	JDBC	32.88	1.2 Billion
6	12GB	2	JDBC	34.77	1.2 Billion

5. No vendor lock-in : Monthly billing on fixed based charges (which will be reducing based on current scenario)

6. Ability to cater to most of the in-house G-level data Quality use-cases.

Thanks

Gaurav agarwal

**From:** Agarwal, Gaurav [Optum]

**Sent:** Wednesday, May 24, 2023 1:12 PM

**To:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Subject:** RE: Can you please compare DQaaS against these products

Hi Dave,

Please find the detailed market research where we did extensive comparison between commercial and open sources products.

Here, is the document viewpoint wrt shortlisted products DQ product being picked for healthcare platform.

[https://dqaas-docs.optum.com/docs/Architecture/PointOfView/dqaas\\_poc\\_details](https://dqaas-docs.optum.com/docs/Architecture/PointOfView/dqaas_poc_details)

**Costing : \$ 8774.11 per month per tenant** which is expected to go down next year looking at the current scenario.

Let me know if you need further details on the same.

Regards

Gaurav agarwal

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Sent:** Wednesday, May 24, 2023 8:18 AM

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>  
**Sent:** Wednesday, May 24, 2023 8:18 AM  
**To:** Agarwal, Gaurav [Optum] <[gaurav\\_agarwal@optum.com](mailto:gaurav_agarwal@optum.com)>  
**Subject:** Can you please compare DQaaS against these products

Gaurav,

Can you please help me fill out the DQaaS capabilities for the features listed in the attached worksheet? Thanks in advance.

Regards,

**Dave Cheema**

(952) 205-0802

RE: Which other group is using DQaaS besides OCDP?



Agarwal, Gaurav [Optum]

To: Cheema, Dave; Ramakrishnan, Anantha  
Cc: Mandal, Aditya; Malaiappan, Nagaraj

Retention Policy UHGIinbox (90 days)

Reply Reply All Forward

Wed 5/24/2023 2:54 AM

Expires 8/22/2023



You replied to this message on 5/24/2023 9:01 AM.

Hi Dave,

Here are other customers who are being engaged with DQaaS

Data lake sources ( 40+ plus)

Optum Clinical Manager

Claim Payment System (UMR)

u-HUB

UHC Global NAM BBDA

Optum Care Delivery

Care management (OCDP)

Common Quality Database : Onboarding

Thanks

Gaurav agarwal

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Sent:** Wednesday, May 24, 2023 7:25 AM

**To:** Agarwal, Gaurav [Optum] <[gaurav\\_agarwal@optum.com](mailto:gaurav_agarwal@optum.com)>; Ramakrishnan, Anantha <[anantha.ramakrishnan@optum.com](mailto:anantha.ramakrishnan@optum.com)>

**Cc:** Mandal, Aditya <[aditya.mandal@optum.com](mailto:aditya.mandal@optum.com)>

**Subject:** Which other group is using DQaaS besides OCDP?

Thank you.

Regards,

**Dave Cheema**

(952) 205-0802



MR on DQ  
for Big Da...



Commercial  
DQ Produ...

RE: Can you please compare DQaaS against these products



Agarwal, Gaurav [Optum]

To: Cheema, Dave

Retention Policy UHGIinbox (90 days)

Reply Reply All Forward

Wed 5/24/2023 2:42 AM

Expires 8/22/2023



You replied to this message on 5/24/2023 9:00 AM.



MR on DQ for Big Data and Streaming Platform.pptx

5 MB



Commercial DQ Product Feature comparison.xlsx



Internal

You replied to this message on 5/24/2023 9:00 AM.



MR on DQ for Big Data and Streaming Platform.pptx  
5 MB



Commercial DQ Product Feature comparison.xlsx  
25 KB

Hi Dave,

Please find the detailed market research where we did extensive comparison between commercial and open sources products.

Here, is the document viewpoint wrt shortlisted products DQ product being picked for healthcare platform.

[https://dqaas-docs.optum.com/docs/Architecture/PointOfView/dqaas\\_poc\\_details](https://dqaas-docs.optum.com/docs/Architecture/PointOfView/dqaas_poc_details)

**Costing : \$ 8774.11 per month per tenant** which is expected to go down next year looking at the current scenario.

Let me know if you need further details on the same.

Regards

Gaurav agarwal

---

**From:** Cheema, Dave <[dave.cheema@optum.com](mailto:dave.cheema@optum.com)>

**Sent:** Wednesday, May 24, 2023 8:18 AM

**To:** Agarwal, Gaurav [Optum] <[gaurav\\_agarwal@optum.com](mailto:gaurav_agarwal@optum.com)>

**Subject:** Can you please compare DQaaS against these products

Gaurav,

Can you please help me fill out the DQaaS capabilities for the features listed in the attached worksheet? Thanks in advance.

Regards,

**Dave Cheema**

(952) 205-0802

# DQaaS Users Experiences

Monday, May 29, 2023 4:39 PM

## Chetan interview questions

How is DQ implemented?

Use DQaaS. Batch, thru Jenkins, service principal. Local deployment, transfer metrics to DQaaS.

What is the level of the data quality - simple/medium/complex? - talk to Param Does Not Know

Is Global ID playing any role for DQ needs? If yes, how?

Every business have their own ASK ID for billing purpose

When does the DQ come into the picture in the data ingestion SDLC?

Multiple places. DQ on destination (Snowflake)

Is it done during streaming, batch, or hybrid?  
batch

If streaming, where is it implemented in the data pipeline? Is it at multiple places or just one place?

Are you using point solution (imperative), commercial product or DQaaS?  
DQaaS

How are monitoring and measurements done, process, tools, etc.? - to Param

How are those metrics presented?

DQ dashboard UI - DQ as a service folks

What are the tools and the languages being used?  
DQ Folks

How is the current process working; any improvement ideas?  
DQ folks. deployment

Metadata - how is data lineage being used?  
Do not know - Param

Is any commercial product used for it?  
NO

Is any other metadata collected/used?  
No

Do you do any data enrichment as a part of the DQ?  
No

+++++

+++++

## Paramjit Nayak questions:

DQ management strategies (process, tool, people)?

DCDM, landed data. Future plan to connect to sources. New markets, sources. 647 DQ rules from workshops. Business quality rules as well.

How is DQ defined for the OCDP platform?

Business and data validation. DQaaS UI

Is it a homegrown solution, any commercial product is used, or DQaaS service is used?

Calliber, home grown solution. Not viable solution. DQaaS is < 10K/month. YAML file for fine tune. Can Group rules together. Minimize duplications

What is the level of the data quality - simple/medium/complex?

3 sets of DQ,

How are monitoring and measurements done, process, tools, etc.?

Volumetrics, dataset, row-level

Do you define any metrics? If yes, which ones?

No

How are those metrics presented?

No

What are the tools and the languages being used?

DQaaS - business define rules

Metadata - how is data lineage being used?

In process

Is any commercial product used for it?

Alaiton is being used for data lineage

Is any other metadata collected/used?

No

Do you do any data enrichment as a part of the DQ?

No. Source SMEs only

DQ you define, is it ingestion/project specific or general?

Is there any reusability of these definitions?

Yes

Where do you get requirements from - business, technology, or empirical?

Business, DQ workshop (took 3-4 months)

What technology and languages are being used?

DQaaS

Where are those definitions persisted, e.g., database, configuration files, etc.

DQaaS

How is the current process working; any improvement ideas?

Working well. Connection to Databricks,

Could you give a brief demo?

Yes

# Deliverables Responses

Monday, May 29, 2023 8:52 PM

DQ implementation plan for TWO focused data quality issues with go-forward processes for proactive issue identification and remediation

## Simple use cases:

- Profiling agent contact data elements for basic DQ dimensions of completeness, format, consistency, etc. For example:
1. Phone numbers – how many formats; included dashes or parentheses; are all numeric, are 10 digits, missing phone, etc.
  2. State codes – Is state code in the State code table; do state codes and zip codes align; missing state codes

## Complex use case:

1. Email (multiple people sharing email address, multiple members have same name and DOB - how do we distinguish?)
2. Multiple address locations - how do we identify the most relevant

**Note:** Agent is looking like a good use case for the DQ tool evaluation and process work. The volumes are manageable, no member PII, and data quality can vary.

Proactive: Simple use cases (Phone numbers, State codes)

Reactive: Complex use cases (Email, Multiple addresses)

## COMPLEX CASE

Reactive – deploy weekly query to monitor COMPAS email address for over-merge

Proactive – add business rule to COMPAS to prevent sharing of same email by two different members of same name.

## SIMPLE CASE

Neither Proactive or Reactive – a new data source for Enterprise Analytics should always be profiled for quality.

I think we have other cases from history docs that we could

List as anticipated use cases

DQ Tool Recommendations to replace Global ID

Rationale to replace Global ID

SDLC does not incorporate data management into lifecycle, struggle to reconcile/implement Global ID

Product to replace Global ID

Early candidate appears to DQaaS

Documented SoPs and implementation plan to incorporate DQ into software development and EA process

Move DQ as far left as pragmatically possible

Implement DQ proactively as pragmatically as possible

Select a DQ tool as reusable as possible

Ensure DQ is a first class citizen

Ensure requirements are as clearly defined and understood as possible

There should be some checkpoints to ensure DQ is factored in and implemented appropriately

Since Data Quality is new to this group, a DQ COE recommended

KPI's to monitor and measure the business and technology adoption of DQ for ONE inflight project??

Metadata lineage for THREE selected use cases for AOP / COMPAS migration. Do we have to implement metadata lineage tool??

# DQ Use case templates

Monday, June 12, 2023 10:36 PM



Use-Case-T  
emplate-05



Use-Case-T  
emplate-06

Template for a data quality use case:

- **Use case name:** This is a brief description of the use case.
- **Business objective:** This is the business goal that the use case is trying to achieve.
- **Data quality requirements:** This is a list of the data quality requirements that must be met in order to achieve the business objective.
- **Data quality metrics:** This is a list of the metrics that will be used to measure the success of the use case.
- **Data quality activities:** This is a list of the activities that will be performed in order to improve the data quality.
- **Data quality resources:** This is a list of the resources that will be needed to implement the use case.
- **Data quality timeline:** This is a timeline for the implementation of the use case.
- **Data quality risks:** This is a list of the risks that could impact the success of the use case.
- **Data quality contingency plans:** This is a list of the plans that will be used to mitigate the risks to the success of the use case.

This template can be used to document any data quality use case. By following this template, you can ensure that your use case is well-defined and that all of the necessary steps are taken to improve the data quality.

Here are some additional tips for writing a data quality use case:

- Be specific. The use case should be specific enough that it can be implemented.
- Be realistic. The use case should be realistic in terms of the resources that are available.
- Be measurable. The use case should have measurable goals and objectives.
- Be flexible. The use case should be flexible enough to adapt to changes in the business environment.
- Be collaborative. The use case should be developed in collaboration with stakeholders from across the organization.

By following these tips, you can write a data quality use case that will help you improve the quality of your data.

# Notes

Sunday, August 6, 2023 9:58 AM

## OAS Checkpoint

Ragesh's team, John, Gaurav, and Sandeep

+++++

AAI AARP Membership API - Execution Planning

Data Separation on Member Portal - Idea Expansion

Plan tables (around 25) all need to hang together

PEOPLE - there are fewer SME's in Fox than we had in UCPS. And those folks are more IT.

Business does not understand the tables.

PROCESS - We don't have SOP for maintaining them (we have SOP's for a few tables like Medicare Constants, CPT codes and ICD codes)

TECHNOLOGY - Basic maintenance screens for Plan tables were not developed in FOX – put on Tech Debt list

FYI:

Plan State Information (PSIF)

Wave 3 - PTIF (Fox\_app\_PDM\_0230106

(Note: not many people understand these tabless)

Note: Even though 'Execution Delivery' is determined too late for DG to be tracking progress, this is still a stage where gaps can occur. For example, in FOX claims, the scope to build maintenance screens and reports for the Plan tables was put on the Tech Debt list

Issue List:

in FOX claims, the scope to build maintenance screens and reports for the Plan tables was put on the Tech Debt list should be put on the Issue list

+++++

PYTHON/PYSPARK - 7/10

SCALA - 7/10

DATABRICKS - 7/10

ADF (Orchestration) - 7/10

COMMUNICATION - 7/10

AZURE ECOSYSTEM(General Exposure) - 5/10

ADLS (should be part of Azure) - 5/10

-----

AIRFLOW (Nice to have) - 3 /10

CICD (General Exposure, not SME) - 3 /10

QUALITYGATE-SECURITY/VULNERABILITIES - 3 /10

Exposure to scans, security issues/coding standards, and remediation

+++++

RAKESH REDDY THUMMALA

Dheeraj Kushakula

Jaswanth Gade

BHARATH KUMAR SUDINI

Shiva Reddy Arrabothu

+++++

Meeting: Feedback from Dinesh 07/28/2023

What expectation in the executive summary

what were the workstreams and what we're delivering  
Message needs to be crisp and deliverables need to be called out

Heading should tell what you're trying to tell them  
Start with the message, e.g., best is contextual  
Change the sections on the tool recommendations slide  
Opportunities section needs to be revised  
Give highlights of DQ enablement  
DQ Operationalization - SDLC Enablement - too much sausage making. Tell them what you need to tell them

need 5-6 very crisp slides  
Anantha and Scott co-present presentation  
Scott should own the next steps, e.g., reusable templates and guidelines  
Push out the meeting with Liv  
David Searfass should be our advocates  
+++++  
One slide --> slide 9 and 10  
DC will talk about the tool  
Key messages: Plandata has no people, process or technology, e.g., High deductible G Data

Complete  
Compile all artifacts

Adoption and Enablement  
+++++  
Meeting 07/31/2023 with Scott and KIM  
Subject: Review feedback  
Hard to see  
Portfolio Management  
PPM - project portfolio Management (Prioritization --> PPM)  
Identify ownership for action -->  
PPR  
Change to Portfolio Management  
DQ Enablement --> Portfolio Management  
8/23, 9/23, 10/23

DQ Tool Replacement, DQ Embedment into SDLC / PDLC Processes, Analysis of Metadata Lineage

+++++  
SKILL LEVEL RATING (1-10, 10 being the highest in proficiency)  
PYTHON/PYSPARK 7/10 - 2  
SCALA 7/10 - 3  
DATABRICKS 7/10 - 0  
ADF (Orchestration) 7/10 - 0  
COMMUNICATION 7/10 - 8  
AZURE ECOSYSTEM(General Exposure) 5/10 - 4  
ADLS (should be part of Azure) 5/10 - 1

-----  
AIRFLOW (Nice to have) 3 /10  
CICD (General Exposure, not SME) 3 /10  
QUALITYGATE-SECURITY/VULNERABILITIES 3 /10

++++++

= $(G6 * (\$E6 + \$F6)) / (\$E\$14 + \$F\$14)$

G6 = OAS rating

$\$E6$  = client rating

$\$F6$  = Indus. Rating

$\$E\$14$  = Sum of client ratings

$\$F\$14$  = Sum of Indus. ratings

= $(\$E\$2 * (\$C\$2 + \$D\$2)) / (\$C\$44 + \$D\$44)$

C44, D44

= $(\$E\$2 * (\$C\$2 + \$D\$2)) / (\$C\$44 + \$D\$44)$

++++++  
Meeting 08/02/2023 with Dinesh and John

Subject: strategy

DC to reachout to Mattilion, Informtica for ETL tools and cost implications

DC to talk to Anantha on Governance

FunctionL capabilities

OAS MRIS Assessment

++++++

Pre-sales

Requirements gathering and analysis

Data modeling (transactional)

Data Security

Data governance, Data services, Crystal & PowerBI reporting

SSIS and SSRS, SQL/PL-SQL

Strategic planning, Roadmaps

Research papers

Adoption architectures

When I start on a project, i know very little about it. But when I'm done with it, there is very little that I don't know about. In fact, I get submerged in it so much that I almost forget everything else

I work hard and I'm quick learner. For example, MRIS DQ project, I knew very little about DQ

Operationalization. But now, I have a pretty solid understanding of it, I'm sure Anantha can attest to it.

---

= $(\$D\$2 * (\$C\$2)) / (\$C\$43)$

C2 = Client rating

D2 = Analyst rating

$\$C\$43$  = Sum of client ratings

++++++

Total pool and things we're tracking (Enterprise DQ Enablement State and DQ Enablement State)

Definitions, Legend, Clear Labels

++++++

Send roles description to Dinish only or Scott Johnson as well?

++++++

SUMPRODUCT(D3:D10, C3:C10)/SUM(C3:C10)

C = weight (importance)

D = rating

++++++

