Leveraging DataOps to build India’s National Data Platform

Prukalpa - Cofounder, Atlan

Subsurface LIVE

The Cloud Data Lake Conference

Hi, I'm Prukalpa 👋

Lifelong data practitioner

Founder of Atlan

Atlan, we are a Github for data teams
Help teams like Postman, Plaid, Unilever, Juniper democratize data

Tons of successes and failures in building data culture

India's national data platform used by Prime Minister
Global SDG monitoring for the United Nations
200+ data projects

We started as a data team ourselves using data science for social good

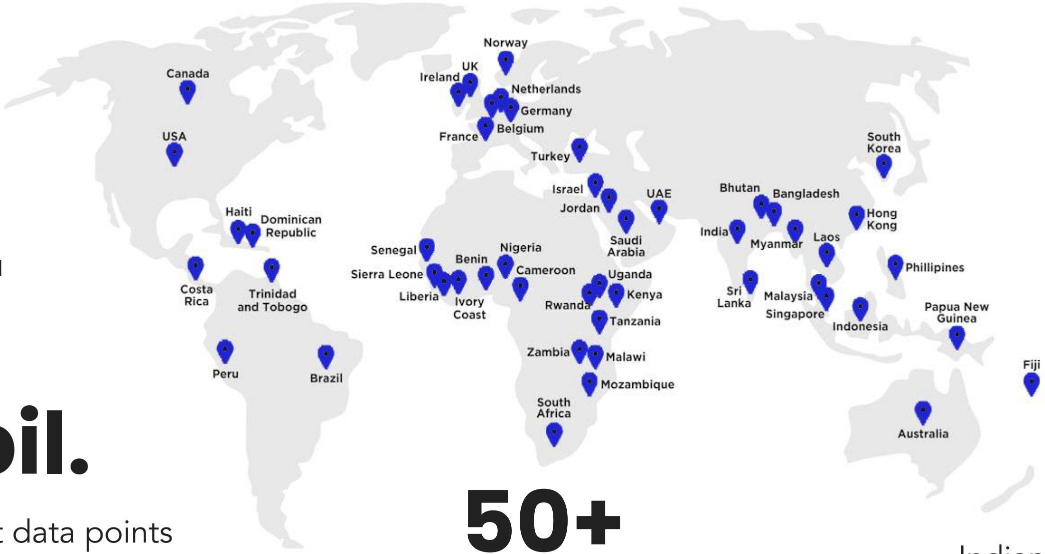
110 bil.

external data points

ingested, cleaned and visualized

6.5 bil.

satellite imagery pixels processed



1.5 bil.

government data points

aggregated in real-time

50+

countries with diverse set of organizations

500 mil.

Indian citizens' data processed

THE WORLD BANK

United Nations

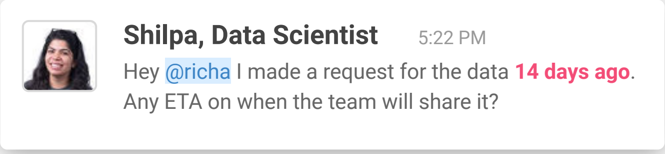
BILL & MELINDA GATES foundation

Government of INDIA

But internally, everyday was chaos.

#team-datascience

Data discovery

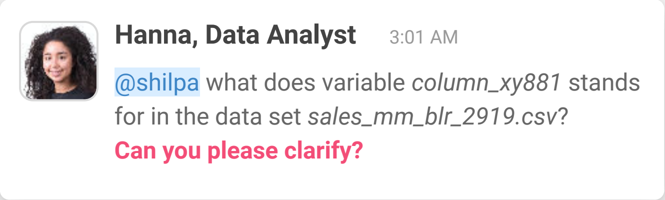


Shilpa, Data Scientist 5:22 PM

Hey @richa I made a request for the data **14 days ago**. Any ETA on when the team will share it?

Human tribal knowledge

Private Chat

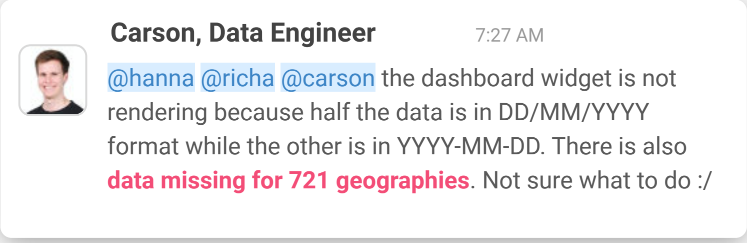


Hanna, Data Analyst 3:01 AM

@shilpa what does variable *column_xy881* stands for in the data set *sales_mm_blr_2919.csv*? **Can you please clarify?**

#team-frontend

Data visibility

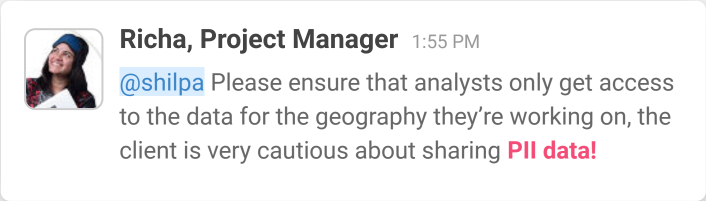


Carson, Data Engineer 7:27 AM

@hanna @richa @carson the dashboard widget is not rendering because half the data is in DD/MM/YYYY format while the other is in YYYY-MM-DD. There is also **data missing for 721 geographies**. Not sure what to do :/

Data governance

#project-gb-data



Richa, Project Manager 1:55 PM

@shilpa Please ensure that analysts only get access to the data for the geography they're working on, the client is very cautious about sharing **PII data!**

IT'S NOT JUST ABOUT IDEAS. IT'S ABOUT MAKING IDEAS HAPPEN
DO IT!

That’s how we started the **assembly line project**

Our team became **6X** more agile.

Building The World's Largest Government Data Lake - DISHA Platform

Built by an 8 member team in 12 months

4 hadn’t pushed a line of code to production before

Here’s the backstory

Our diverse **Humans of Data**

Hannah the **Analyst**

python R

Shilpa the **Data Scientist**

jupyter Paxata colab

Jessie the **Compliance Lead**

okta X M

Prukalpa the **CDO**

tableau

Richa the **Project Manager**

M X

Carson the **Data Visualizer**

tableau Google Data Studio

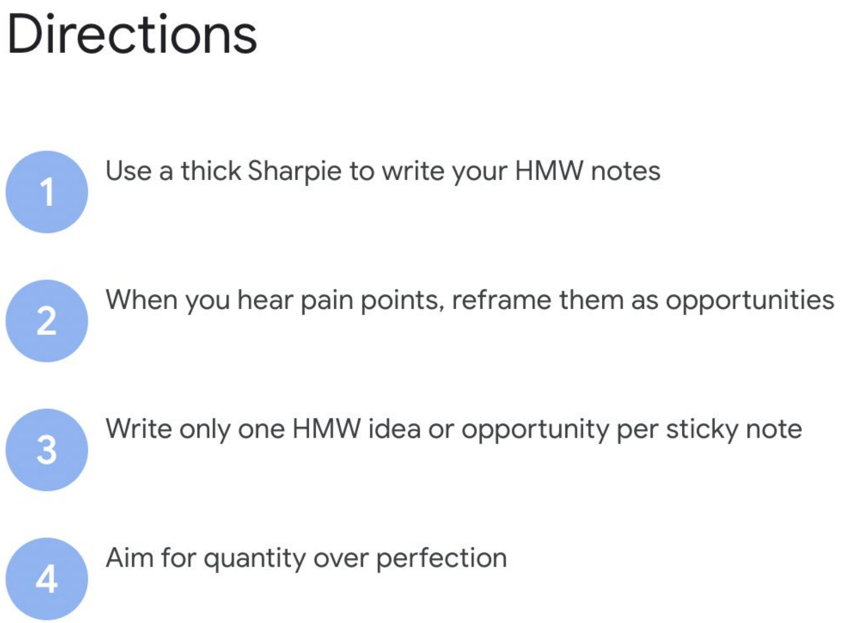
Patrick the **Data Engineer**

Airflow mongoDB

Christine the **Consultant**

X T G

Google Design Sprint
“HMW” How Might We exercise



Directions

1. Use a thick Sharpie to write your HMW notes
2. When you hear pain points, reframe them as opportunities
3. Write only one HMW idea or opportunity per sticky note
4. Aim for quantity over perfection

https://designsprintkit.withgoogle.com/methodology/phase1-understand/hmw-sharing-and-affinity-mapping

HMW Create a High Performance Culture?

HMW Plan timelines better?

HMW reduce repetitive tasks?

HMW reduce scope creep with customers?

HMW Build more trust in our data and improve data quality?

HMW leverage our learnings across projects better?

HMW ensure high quality output without variation across different individuals?

HMW improve collaboration between data engineering & analysts?

HMW onboard new analysts faster and better?

HMW ensure we are solving the problem the best way instead of meeting deadlines?

HMW reduce dependency on individuals? HIT BY THE BUS syndrome.

HMW reduce dependencies on engineering?

HMW Prioritise tasks across projects?

HMW reduce troubleshooting time?

Our Team Charter

High Performance Culture

Can we ensure we meet our deadlines/ targets? How can we plan better? How can we balance this with a culture of experimentation and innovation?

Ecosystem of Trust

Our team will always be small and diverse. How can we ensure that we create a strong ecosystem of trust where diverse people: engineers, analysts, scientists etc trust each other?

Reduce Repetitive Tasks

Create automation, reusability and reproducibility in order to reduce repetitive tasks and improve data team productivity.

Guaranteed Quality of Outputs

How can we ensure that irrespective of the person /people involved in a project, the output will be high quality and something that we would be proud of?

Create a resilient team

“Hit by the bus” syndrome. Never be in a situation where an analyst leaving could endanger the promises we made to our customers.

The two drivers in our journey to becoming 6X more agile

Data stack

Culture stack

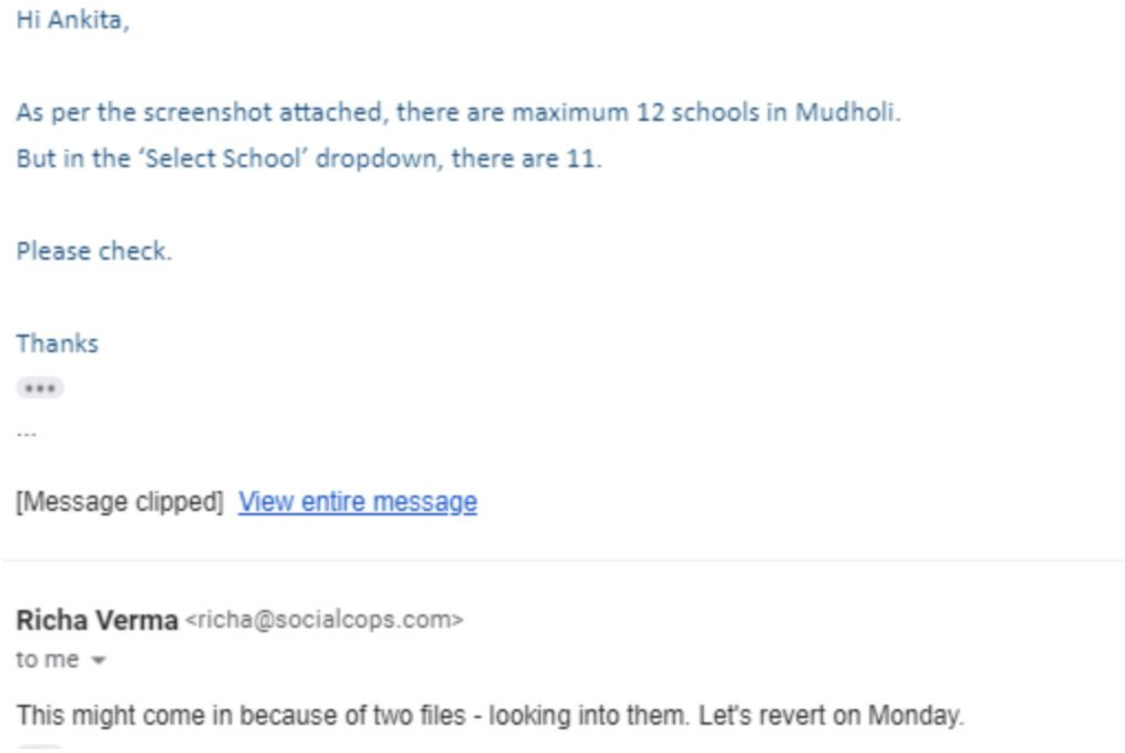
The Modern Data Stack

The Modern Data Culture Stack

1.

The “data stack” that powered our human stack

OK, THAT NUMBER DOESN’T LOOK RIGHT.



Hi Ankita,

As per the screenshot attached, there are maximum 12 schools in Mudholi.

But in the ‘Select School’ dropdown, there are 11.

Please check.

Thanks

...

[Message clipped] View entire message

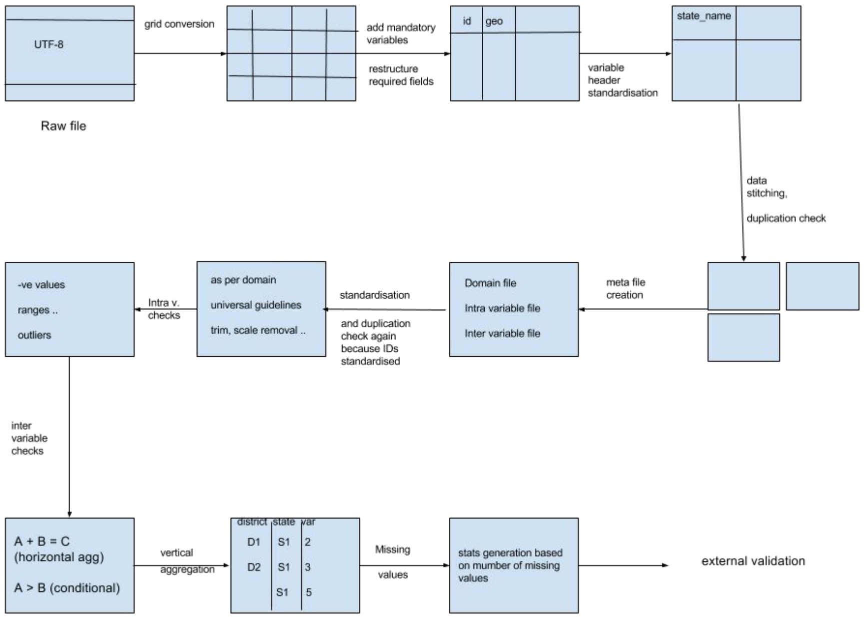
Richa Verma <richa@socialcops.com>

to me

This might come in because of two files - looking into them. Let's revert on Monday.

HMW ensure that our data & insights are actually accurate?

HMW ensure all our analysts follow standard data cleaning processes?



UTF-8

Raw file

grid conversion

add mandatory variables

restructure required fields

id	geo

variable header standardisation

state_name

data stitching,

duplication check

meta file creation

Domain file

Intra variable file

Inter variable file

standardisation

and duplication check again because IDs standardised

as per domain

universal guidelines

trim, scale removal ..

Intra v. checks

-ve values

ranges ..

outliers

inter variable checks

A + B = C (horizontal agg)

A > B (conditional)

vertical aggregation

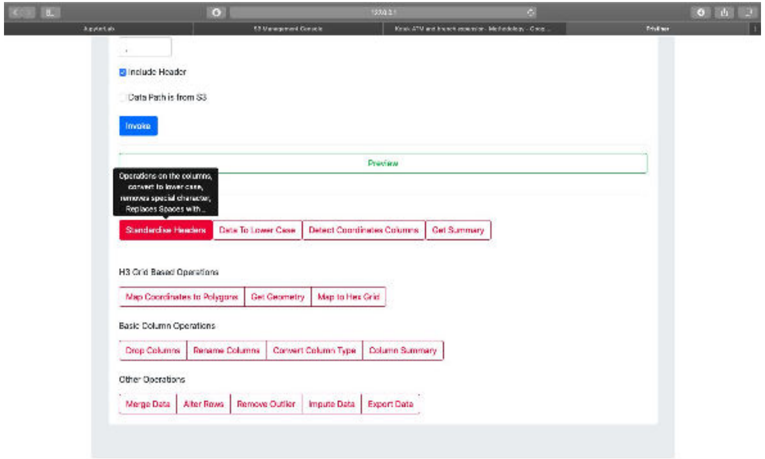
district	state	var
D1	S1	2
D2	S1	3
	S1	5

Missing values

stats generation based on number of missing values

external validation

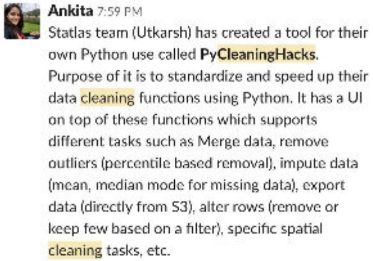
SocialCops framework data cleaning & checking process



☑ Include Header

☐ Data Path is from S3

Operations on the columns, convert to lower case, removes special character, Replaces Spaces with...



Ankita 7:59 PM

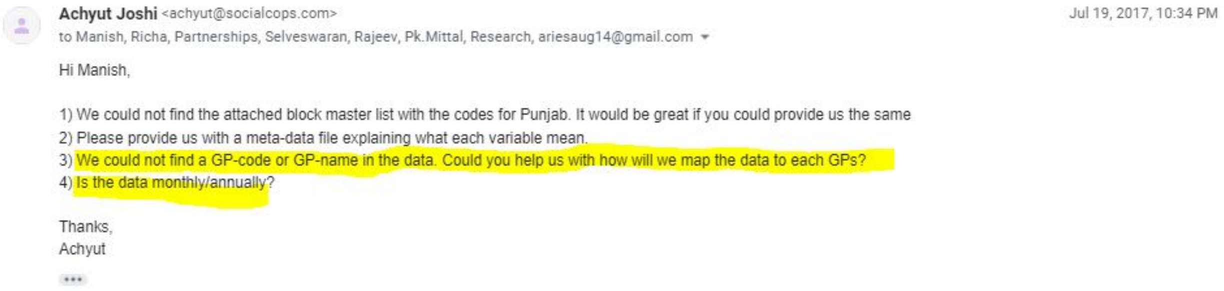
Statlas team (Utkarsh) has created a tool for their own Python use called **PyCleaningHacks**. Purpose of it is to standardize and speed up their data cleaning functions using Python. It has a UI on top of these functions which supports different tasks such as Merge data, remove outliers (percentile based removal), impute data (mean, median mode for missing data), export data (directly from S3), alter rows (remove or keep few based on a filter), specific spatial cleaning tasks, etc.

Analytics Engineering & Data Quality Testing

dbt

great_expectations

THE (META) DATA CONTEXT PROBLEM



Achyut Joshi <achyut@socialcops.com>

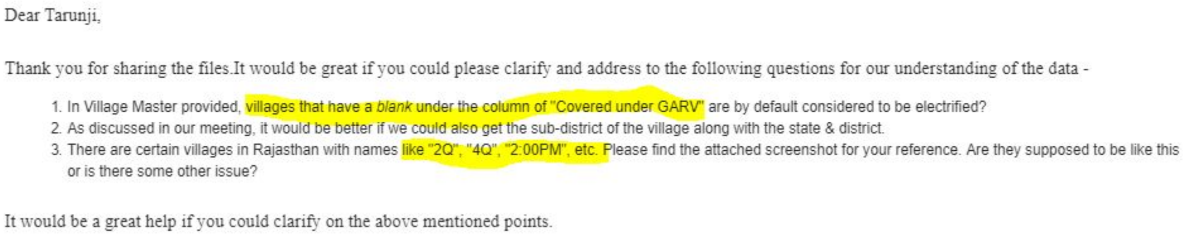
Jul 19, 2017, 10:34 PM

to Manish, Richa, Partnerships, Selveswaran, Rajeev, Pk.Mittal, Research, ariesaug14@gmail.com

Hi Manish,

1) We could not find the attached block master list with the codes for Punjab. It would be great if you could provide us the same
2) Please provide us with a meta-data file explaining what each variable mean.
3) We could not find a GP-code or GP-name in the data. Could you help us with how will we map the data to each GPs?
4) Is the data monthly/annually?

Thanks,
Achyut



Dear Tarunji,

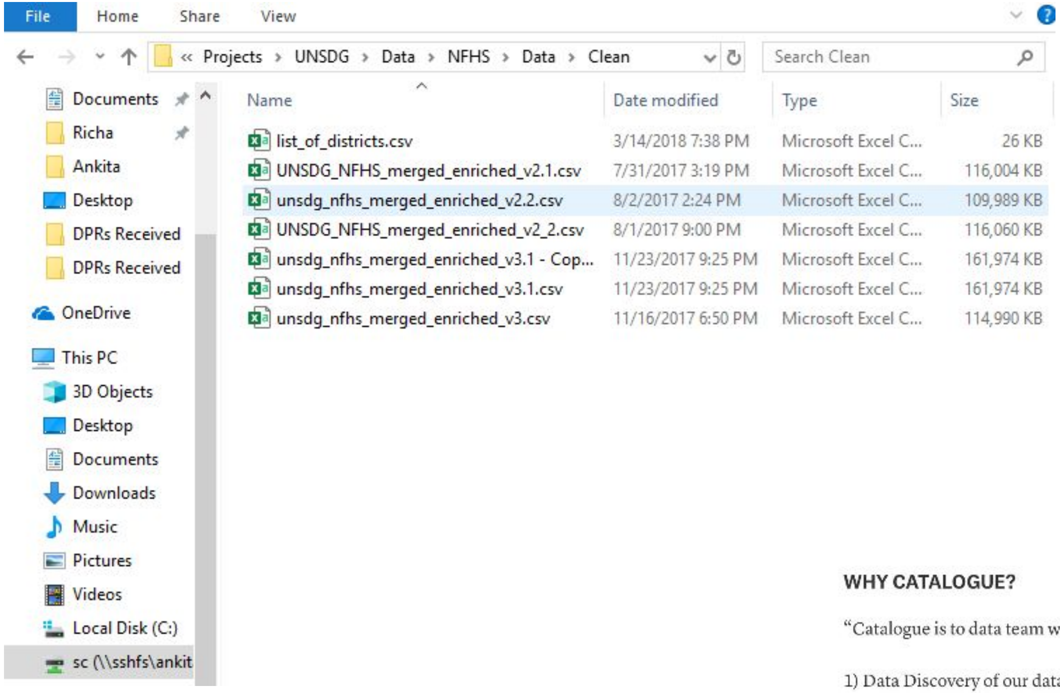
Thank you for sharing the files.It would be great if you could please clarify and address to the following questions for our understanding of the data -

1. In Village Master provided, villages that have a *blank* under the column of "Covered under GARV" are by default considered to be electrified?
2. As discussed in our meeting, it would be better if we could also get the sub-district of the village along with the state & district.
3. There are certain villages in Rajasthan with names like "2Q", "4Q", "2:00PM", etc. Please find the attached screenshot for your reference. Are they supposed to be like this or is there some other issue?

It would be a great help if you could clarify on the above mentioned points.

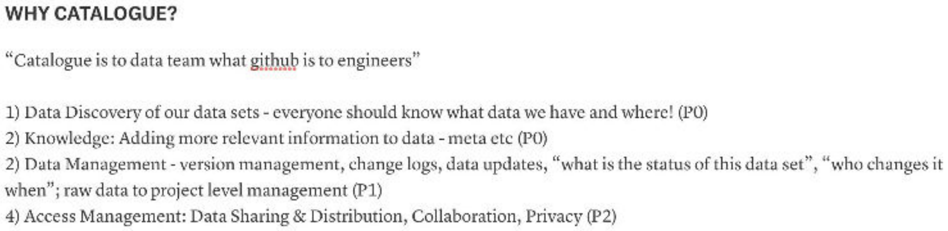
HMW ensure 100% context before we start working on a dataset?

UH... WHICH VERSION?



Name	Date modified	Type	Size
list_of_districts.csv	3/14/2018 7:38 PM	Microsoft Excel C...	26 KB
UNSDG_NFHS_merged_enriched_v2.1.csv	7/31/2017 3:19 PM	Microsoft Excel C...	116,004 KB
unsdg_nfhs_merged_enriched_v2.2.csv	8/2/2017 2:24 PM	Microsoft Excel C...	109,989 KB
UNSDG_NFHS_merged_enriched_v2_2.csv	8/1/2017 9:00 PM	Microsoft Excel C...	116,060 KB
unsdg_nfhs_merged_enriched_v3.1 - Cop...	11/23/2017 9:25 PM	Microsoft Excel C...	161,974 KB
unsdg_nfhs_merged_enriched_v3.1.csv	11/23/2017 9:25 PM	Microsoft Excel C...	161,974 KB
unsdg_nfhs_merged_enriched_v3.csv	11/16/2017 6:50 PM	Microsoft Excel C...	114,990 KB

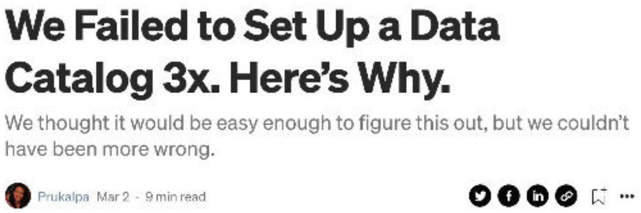
HMW ensure we use the right dataset and know the difference between versions?



WHY CATALOGUE?

"Catalogue is to data team what github is to engineers"

1) Data Discovery of our data sets - everyone should know what data we have and where! (P0)
2) Knowledge: Adding more relevant information to data - meta etc (P0)
2) Data Management - version management, change logs, data updates, "what is the status of this data set", "who changes it when"; raw data to project level management (P1)
4) Access Management: Data Sharing & Distribution, Collaboration, Privacy (P2)



We Failed to Set Up a Data Catalog 3x. Here's Why.

We thought it would be easy enough to figure this out, but we couldn't have been more wrong.

Prukalpa Mar 2 · 9 min read

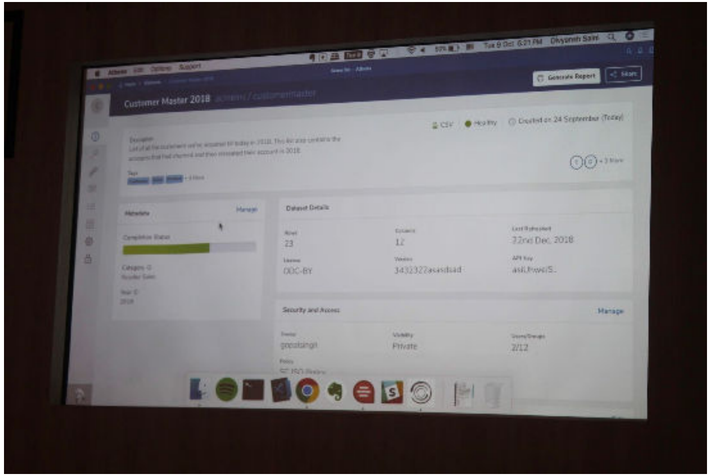
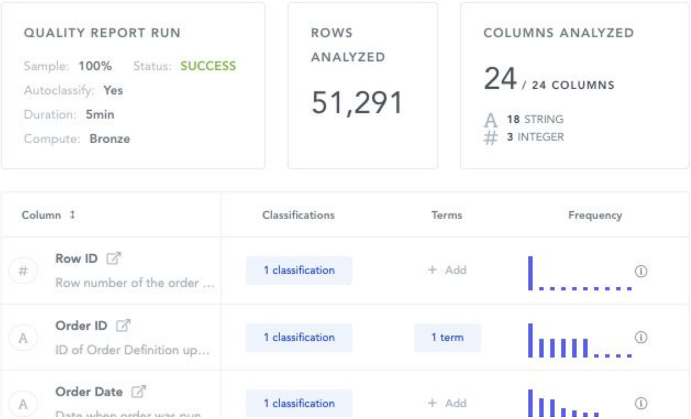


Image: First version of Atlan's data workspace



QUALITY REPORT RUN

Sample: 100% Status: SUCCESS
Autoclassify: Yes
Duration: 5min
Compute: Bronze

ROWS ANALYZED

51,291

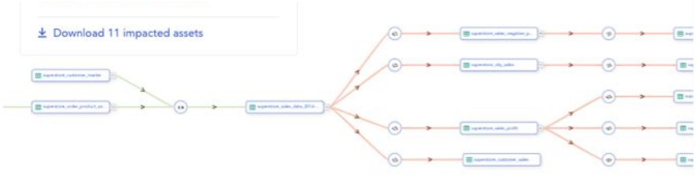
COLUMNS ANALYZED

24 / 24 COLUMNS

A 18 STRING
3 INTEGER

Column ↕	Classifications	Terms	Frequency
# Row ID Row number of the order ...	1 classification	+ Add	
A Order ID ID of Order Definition up...	1 classification	1 term	
A Order Date Date when order was pun	1 classification	+ Add	

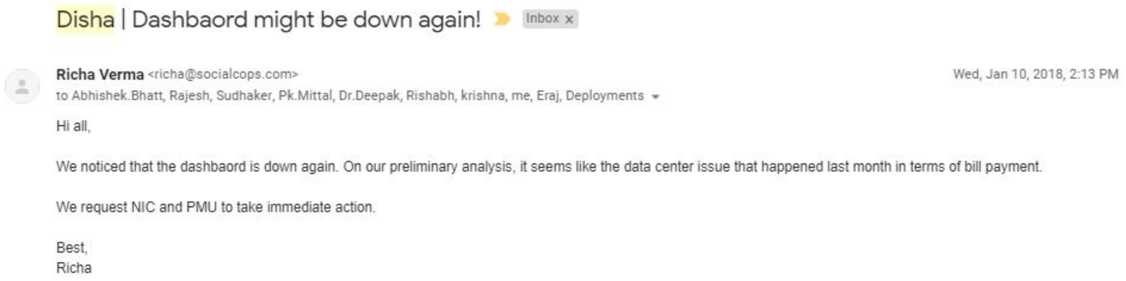
Image: Our first internal tool auto generate a data profiling report to answer all the open questions we had about the data (Frequency distribution Histogram, Column Types, Missing Values, Outliers)



Download 11 impacted assets

Active Metadata Platform (Data Catalog 3.0): atlan

THE BROKEN DASHBOARD



Disha | Dashbaord might be down again!

Richa Verma <richa@socialcops.com>

Wed, Jan 10, 2018, 2:13 PM

to Abhishek.Bhatt, Rajesh, Sudhaker, Pk.Mittal, Dr.Deepak, Rishabh, krishna, me, Eraj, Deployments

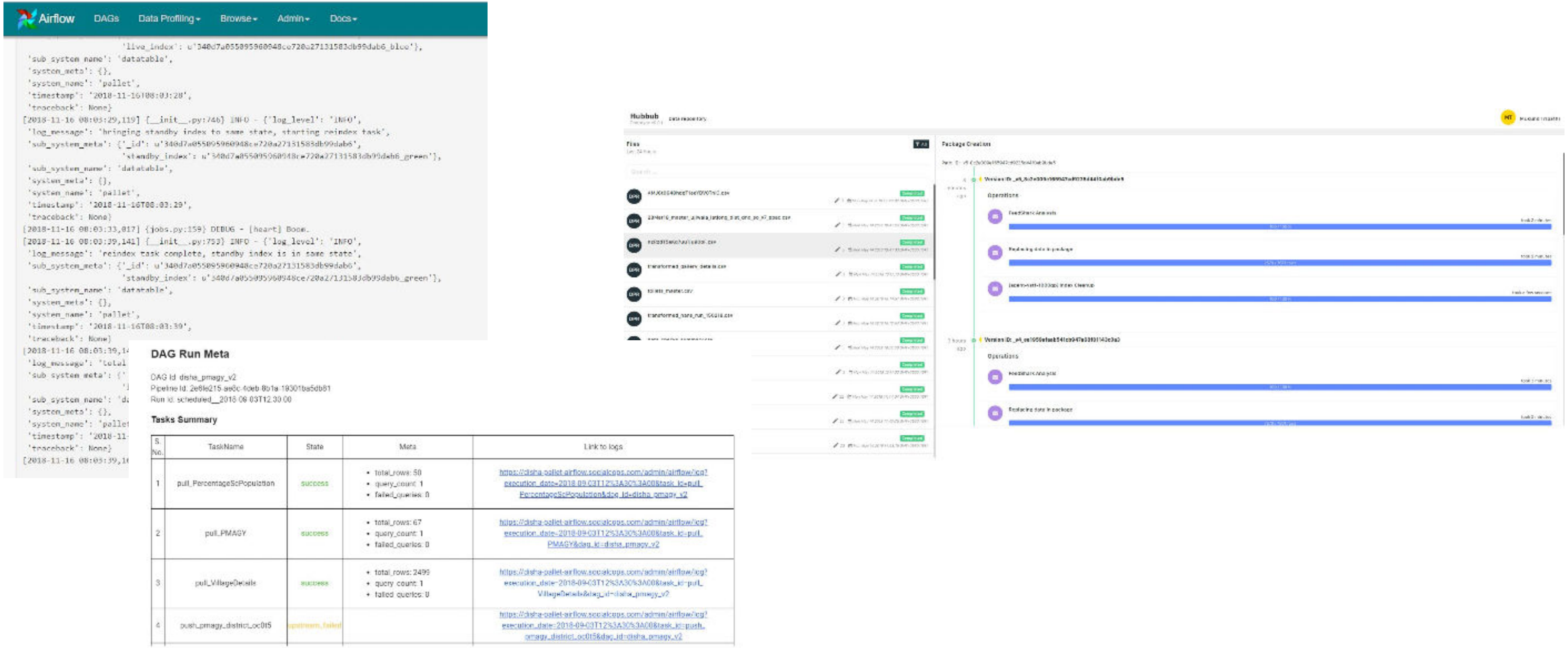
Hi all,

We noticed that the dashbaord is down again. On our preliminary analysis, it seems like the data center issue that happened last month in terms of bill payment.

We request NIC and PMU to take immediate action.

Best,
Richa

HMW know if a dashboard breaks or something goes wrong before a client does?



```
                       'live_index': u'340d7a055095960948ce720a27131583db99dab6_blue'},
 'sub system name': 'datatable',
 'system_meta': {},
 'system_name': 'pallet',
 'timestamp': '2018-11-16T08:03:28',
 'traceback': None}
[2018-11-16 08:03:29,119] {__init__.py:746} INFO - {'log_level': 'INFO',
 'log_message': 'bringing standby index to same state, starting reindex task',
 'sub_system_meta': {'_id': u'340d7a055095960948ce720a27131583db99dab6',
                       'standby_index': u'340d7a055095960948ce720a27131583db99dab6_green'},
 'sub_system_name': 'datatable',
 'system_meta': {},
 'system_name': 'pallet',
 'timestamp': '2018-11-16T08:03:29',
 'traceback': None}
[2018-11-16 08:03:33,017] {jobs.py:159} DEBUG - [heart] Boom.
[2018-11-16 08:03:39,141] {__init__.py:753} INFO - {'log_level': 'INFO',
 'log_message': 'reindex task complete, standby index is in same state',
 'sub_system_meta': {'_id': u'340d7a055095960948ce720a27131583db99dab6',
                       'standby_index': u'340d7a055095960948ce720a27131583db99dab6_green'},
 'sub_system_name': 'datatable',
 'system_meta': {},
 'system_name': 'pallet',
 'timestamp': '2018-11-16T08:03:39',
 'traceback': None}
[2018-11-16 08:03:39,14
 'log_message': 'total
 'sub system meta': {'
                       '
 'sub system name': 'da
 'system_meta': {},
 'system_name': 'pallet
 'timestamp': '2018-11-
 'traceback': None}
[2018-11-16 08:03:39,16
```

DAG Run Meta

DAG Id: disha_pmagy_v2
Pipeline Id: 2e8le215-ae8c-4deb-8b1a-18301ba5db81
Run Id: scheduled__2018-09-03T12:30:00

Tasks Summary

S. No.	TaskName	State	Meta	Link to logs
1	pull_PercentageScPopulation	success	• total_rows: 50 • query_count: 1 • failed_queries: 0	https://disha-pallet-airflow.socialcops.com/admin/airflow/log?execution_date=2018-09-03T12%3A30%3A00&task_id=pull_PercentageScPopulation&dag_id=disha_pmagy_v2
2	pull_PMAGY	success	• total_rows: 67 • query_count: 1 • failed_queries: 0	https://disha-pallet-airflow.socialcops.com/admin/airflow/log?execution_date=2018-09-03T12%3A30%3A00&task_id=pull_PMAGY&dag_id=disha_pmagy_v2
3	pull_VillageDetails	success	• total_rows: 2499 • query_count: 1 • failed_queries: 0	https://disha-pallet-airflow.socialcops.com/admin/airflow/log?execution_date=2018-09-03T12%3A30%3A00&task_id=pull_VillageDetails&dag_id=disha_pmagy_v2
4	push_pmagy_district_oc015	upstream_failed		https://disha-pallet-airflow.socialcops.com/admin/airflow/log?execution_date=2018-09-03T12%3A30%3A00&task_id=push_pmagy_district_oc015&dag_id=disha_pmagy_v2

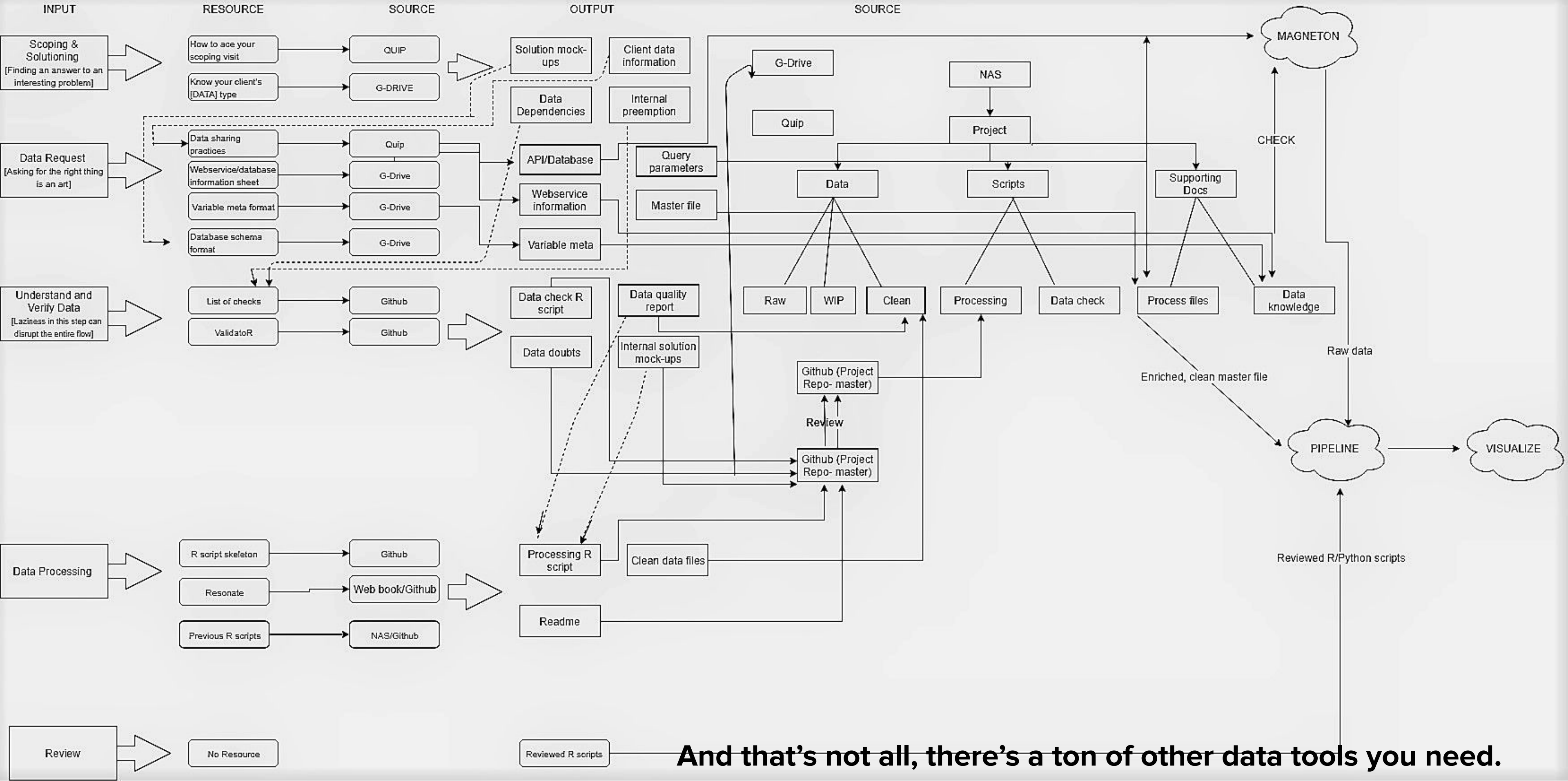
Email Alert

Data Observability Tools:

acceldata

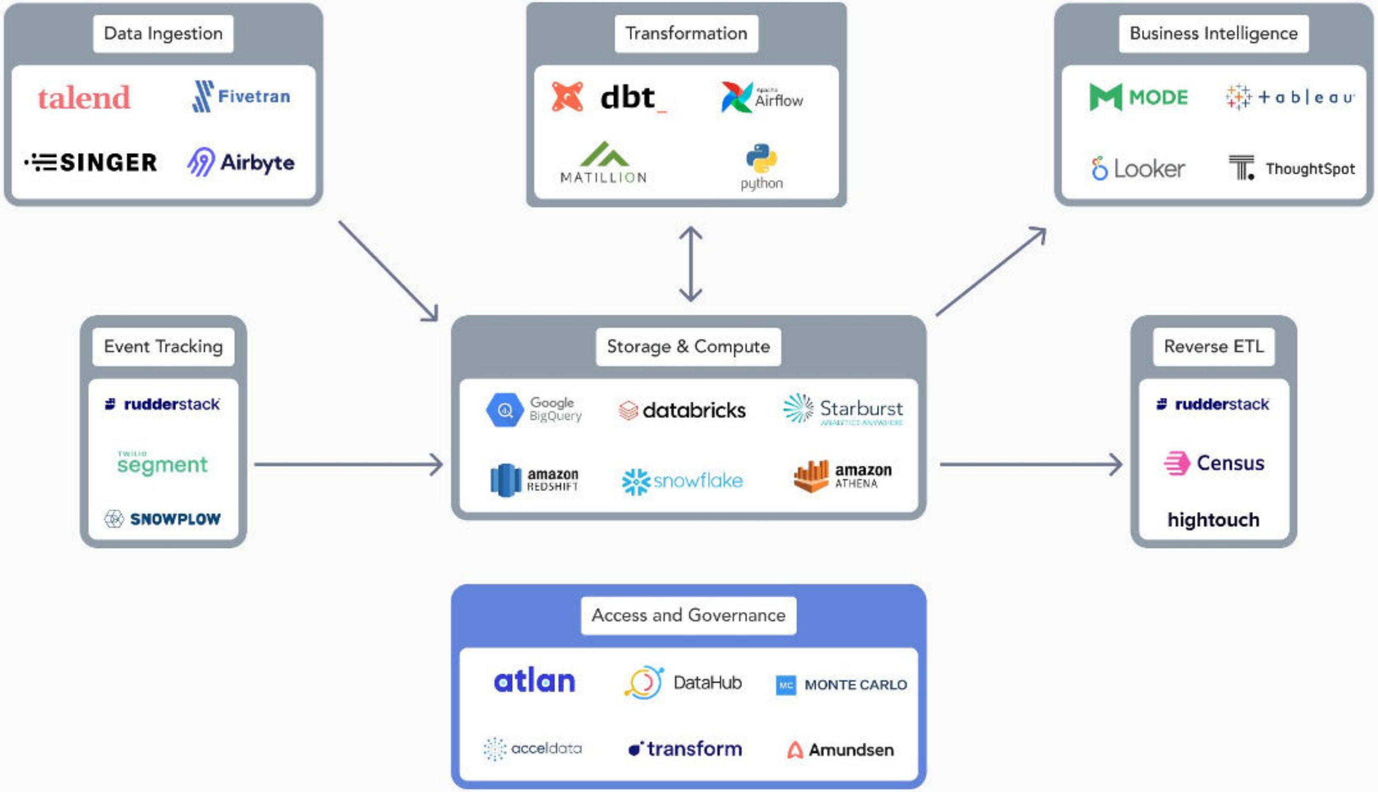
MC MONTE CARLO

Anomalo



And that’s not all, there’s a ton of other data tools you need. TL;DR: Luckily the Modern Data Stack probably has already built it for you.

The modern data stack



Data Ingestion

talend

Fivetran

SINGER

Airbyte

Transformation

dbt_

Airflow

MATILLION

python

Business Intelligence

MODE

tableau

Looker

ThoughtSpot

Event Tracking

rudderstack

segment

SNOWPLOW

Storage & Compute

Google BigQuery

databricks

Starburst

amazon REDSHIFT

snowflake

amazon ATHENA

Reverse ETL

rudderstack

Census

hightouch

Access and Governance

atlan

DataHub

MONTE CARLO

acceldata

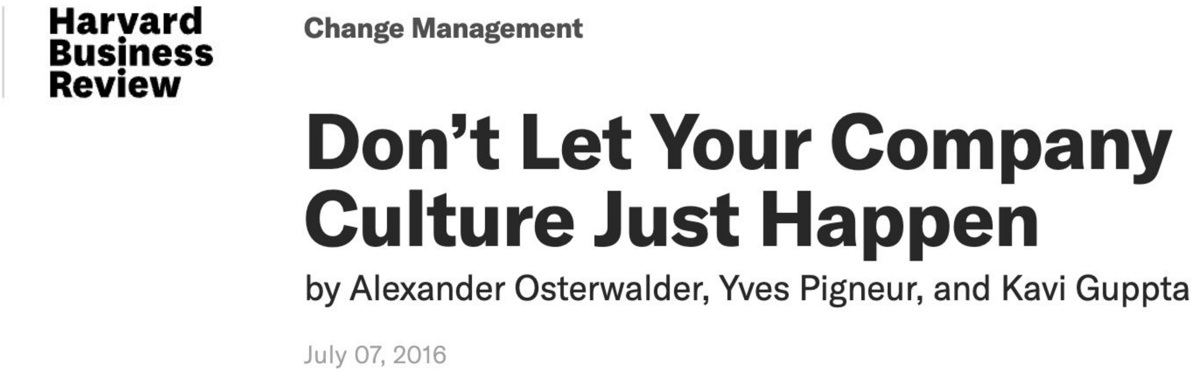
transform

Amundsen

2.

The culture stack that powered our human stack

We need to start talking about the Modern Data Culture Stack



Harvard Business Review

Change Management

Don’t Let Your Company Culture Just Happen

by Alexander Osterwalder, Yves Pigneur, and Kavi Guppta

July 07, 2016

Values → Rituals

Values	Agility	Trust	Collaboration	Innovation
Rituals	Scrum / Agile OKRs Friday Demos Data Product Roadmap	Doc Hours Start/ Stop/ Continue Data F*** Up night Reflection Docs	Colab hours Daily Standups Dependency workflows	Data Brain Trusts Data Product Mindset Google Design Sprints Hackathons

Agility

How do we ensure we plan well and meet our goals effectively?

HMW Create a High Performance Culture?

HMW Plan timelines better?

HMW reduce context switching time?

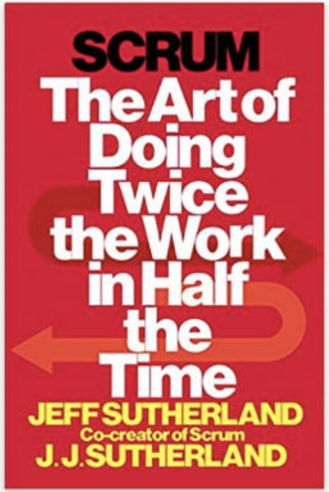
Agility

HMW Create a High Performance Culture?

HMW Plan timelines better?

HMW reduce context switching time?

How do we ensure we plan well and meet our goals effectively?



Agility

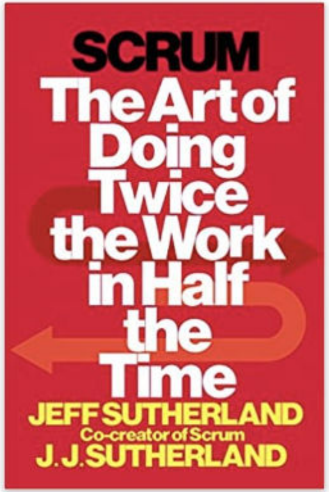
HMW Create a High Performance Culture?

✅

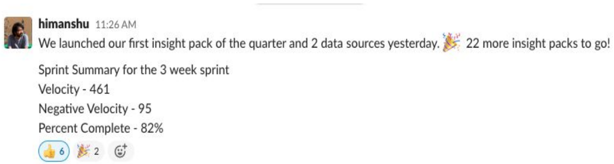
HMW Plan timelines better?

HMW reduce context switching time?

How do we ensure we plan well and meet our goals effectively?



Agile in Data Science improved our velocity 4X



himanshu 11:26 AM

We launched our first insight pack of the quarter and 2 data sources yesterday. 🎉 22 more insight packs to go!

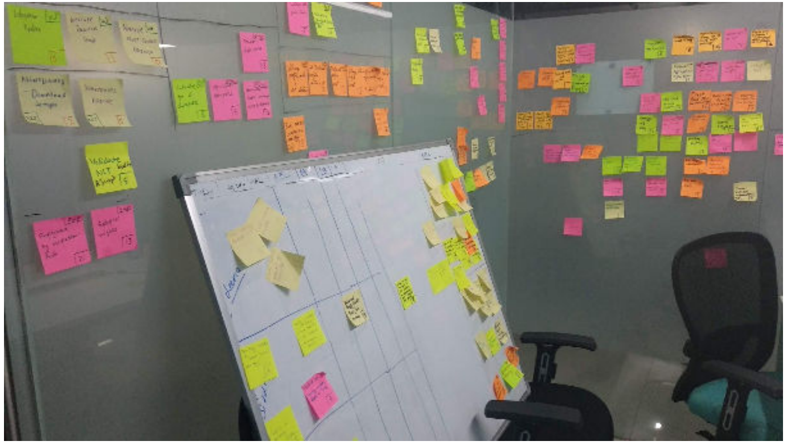
Sprint Summary for the 3 week sprint

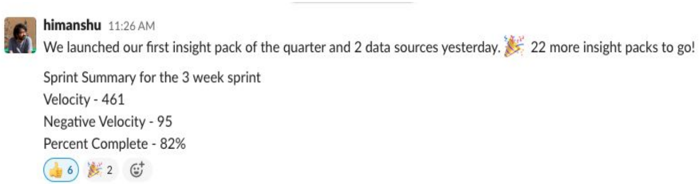
Velocity - 461

Negative Velocity - 95

Percent Complete - 82%

👍 6 🎉 2





himanshu 11:26 AM

We launched our first insight pack of the quarter and 2 data sources yesterday. 🎉 22 more insight packs to go!

Sprint Summary for the 3 week sprint
Velocity - 461
Negative Velocity - 95
Percent Complete - 82%

👍 6 🎉 2

Bottom up, not top down

Everyone on the team read Scrum, we ran internal “learning” sessions before agreeing to “experiment” with it that quarter

Understand and agree on principles

Context Switching principles, “Estimation” of effort, Dependencies

Rituals Matter

Monday planning sessions, Daily standup

The culture of helpful “questioning”

Once we had established principles, our team started asking each other every day: what went wrong in our planning that made it hard for us to achieve our weekly goals/ giving feedback of estimates.

Measure!

Weekly velocity measures and percentage completion goals actually drove us forward

Agility

HMW Create a High Performance Culture?

✅ HMW Plan timelines better?

✅ HMW reduce context switching time?

How do reduce context switching time?

❗❗ Our senior analysts were spending over 3-4 hours a day in context switching time due to interruptions from other team members!

Agility

HMW Create a High Performance Culture?

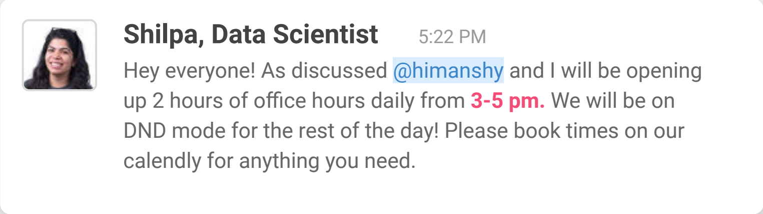
✅ HMW Plan timelines better?

✅ HMW reduce context switching time?

How do reduce context switching time?

❗❗ Our senior analysts were spending over 3-4 hours a day in context switching time due to interruptions from other team members!

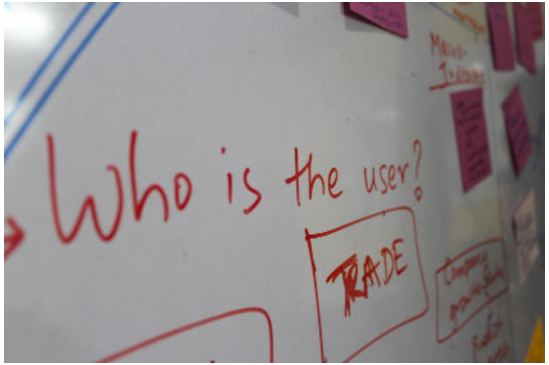
✅ Daily 2 hour Collaboration Hours & Office Hours

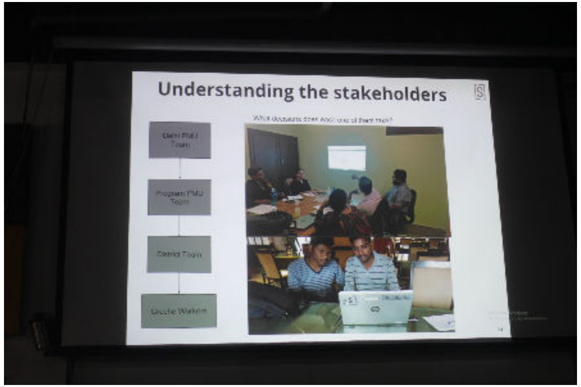


Shilpa, Data Scientist 5:22 PM

Hey everyone! As discussed @himanshy and I will be opening up 2 hours of office hours daily from **3-5 pm.** We will be on DND mode for the rest of the day! Please book times on our calendly for anything you need.

One of our biggest deltas in agility came because of a DNA shift from a “data services” to a “data product” mindset

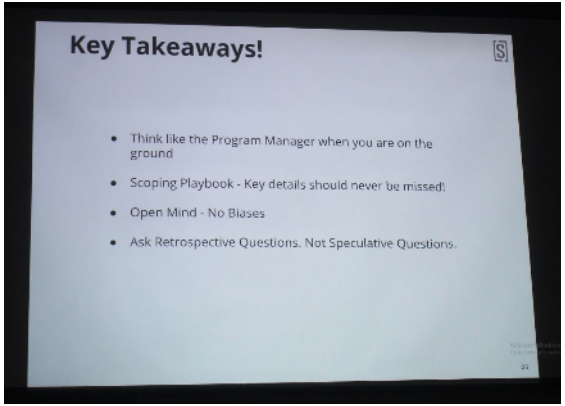




Understanding the stakeholders

What decisions does each one of them take?

Client PMU Team
Program PMU Team
District Team
Creche Worker



Key Takeaways!

- Think like the Program Manager when you are on the ground
- Scoping Playbook - Key details should never be missed!
- Open Mind - No Biases
- Ask Retrospective Questions. Not Speculative Questions.

	Data Services	Data Product
Success Criteria	Successful implementation i.e. did we deliver on time	Successful usage i.e it solve the problem for users & do they use it regularly
Reusability	Single use: Build once, for use by one client	Scalability / Reusability: Build once, for use by many
Requirements / Scoping	Build what the customer asks you to build	Understand commonalities in problems across the customer base & build accordingly

Define Data Product “shipping standards” into your metadata platform

5W1H framework

WHAT

What is the data asset about?

- Table descriptions
- Column descriptions
- Keywords / Glossary Terms

WHY

Why does the data asset exist?

- Data source
- Lineage

WHERE

What is the coverage of the data?

- Coverage
- Language

WHO

Who is responsible/impacted?

- Owners
- Experts

WHEN

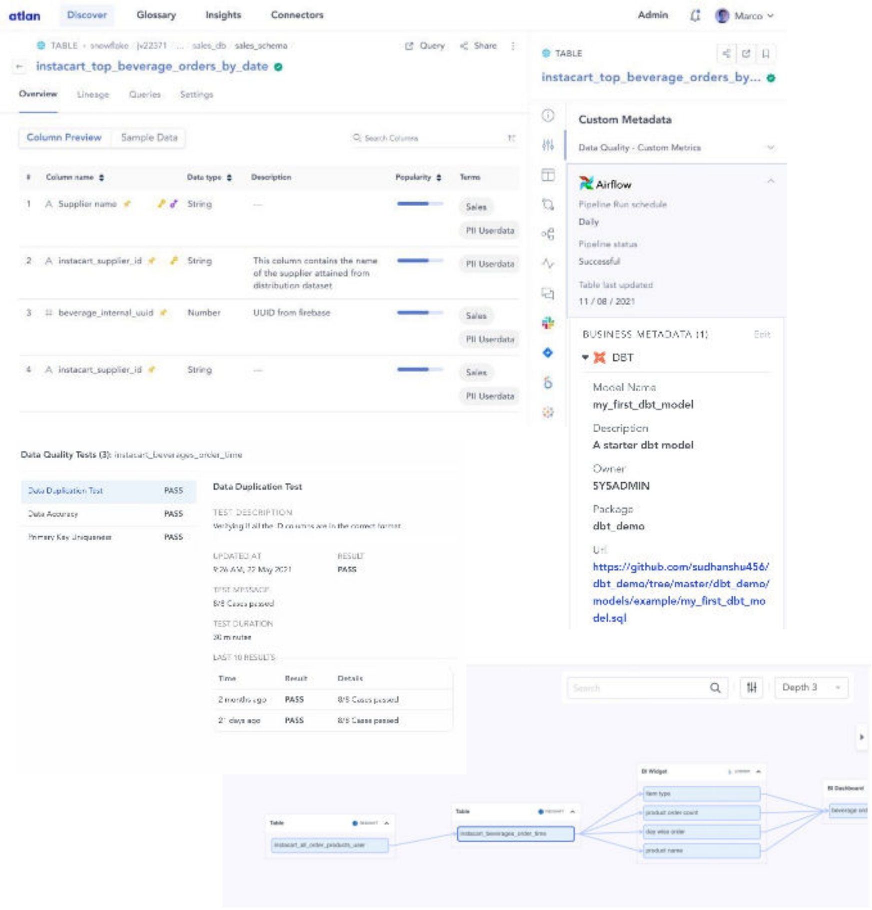
When is it created/updated?

- Update Frequency
- Last Updated
- Quality Metrics

HOW

How can the data asset be used?

- Classification/License
- Use-cases



instacart_top_beverage_orders_by_date

#	Column name	Data type	Description	Popularity	Terms
1	Supplier name	String	—		Sales PII Userdata
2	instacart_supplier_id	String	This column contains the name of the supplier attained from distribution dataset		PII Userdata
3	beverage_internal_uuid	Number	UUID from firebase		Sales PII Userdata
4	instacart_supplier_id	String	—		Sales PII Userdata

Data Quality Tests (3): instacart_beverages_order_time

Data Duplication Test	PASS
Data Accuracy	PASS
Primary Key Uniqueness	PASS

Data Duplication Test

TEST DESCRIPTION
Verifying if all the ID columns are in the correct format

UPDATED AT
9:26 AM, 22 May 2021

RESULT
PASS

TEST MESSAGE
8/8 Cases passed

TEST DURATION
30 minutes

LAST 10 RESULTS

Time	Result	Details
2 months ago	PASS	8/8 Cases passed
21 days ago	PASS	8/8 Cases passed

instacart_top_beverage_orders_by...

Custom Metadata

Data Quality - Custom Metrics

Airflow

Pipeline Run schedule
Daily

Pipeline status
Successful

Table last updated
11 / 08 / 2021

BUSINESS METADATA (1) Edit

DBT

Model Name
my_first_dbt_model

Description
A starter dbt model

Owner
SYSADMIN

Package
dbt_demo

Url
https://github.com/sudhanshu456/dbt_demo/tree/master/dbt_demo/models/example/my_first_dbt_model.sql

Incorporating "human driven" shipping standards into the product shipping process

Data Product Roadmap

Doc Hours

Daily Standups

Creating a weekly "shipping mindset" with Friday Demos

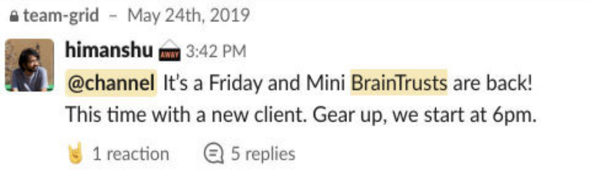
☑ Set the quality standard with the first few internal demos

☑ Make it an event that the team looks forward to

☑ Introducing "showdowns" i.e. best demo of the week can be interesting ways to create healthy competition (ensure that people don't ship bad products just to win Showdown!)

Innovation

HMW ensure we are solving the problem the best way instead of meeting deadlines?

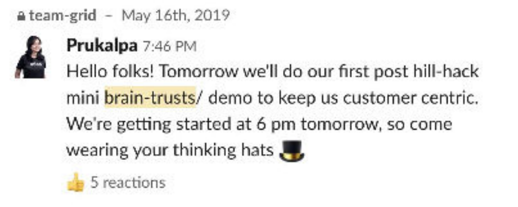


🔒team-grid - May 24th, 2019

himanshu 3:42 PM

@channel It's a Friday and Mini BrainTrusts are back! This time with a new client. Gear up, we start at 6pm.

🤘 1 reaction   5 replies



🔒team-grid - May 16th, 2019

Prukalpa 7:46 PM

Hello folks! Tomorrow we'll do our first post hill-hack mini brain-trusts/ demo to keep us customer centric. We're getting started at 6 pm tomorrow, so come wearing your thinking hats 🎩

👍 5 reactions

✅ The Data Brain Trust Format

03-12-14 | LESSONS LEARNED

Inside The Pixar Braintrust

In this exclusive excerpt from *Creativity, Inc.*, Ed Catmull unveils one of his key management tools–the Pixar Braintrust, which has helped the animation powerhouse score 14 box office hits in a row.

☑ The Briefing: 5 mins. The Business, & problems they want to solve
☑ The Balcony: 5 mins. Group discusses what they understood "I heard X say". The facilitator can only listen & can't speak.
☑ Clarify: 5 mins: The facilitator clarifies the groups understanding
☑ Key Needs: 2 mins: Outlines key needs
☑ Q&A: 15 mins: Think, Pair, Share format
☑ Individual Brainstorm: 5 mins: Ideas go on board
☑ Quiet Reading & Voting on top 5 ideas: 5 mins
☑ Open Discussion

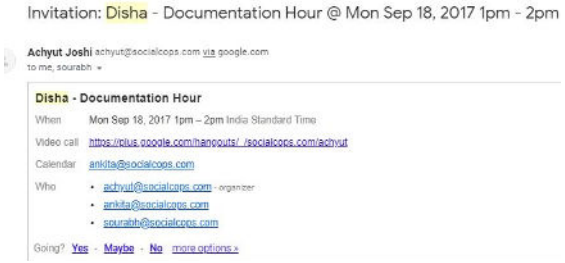
Building Trust & Collaboration

HMW Build a culture of constant improvement & growth

HMW reduce dependency on individuals? HIT BY THE BUS syndrome.

HMW leverage our learnings across projects better?

✅ Setting time aside for reflection & documentation



Invitation: Disha - Documentation Hour @ Mon Sep 18, 2017 1pm - 2pm

Achyut Joshi achyut@socialcops.com via google.com
to me, sourabh

Disha - Documentation Hour

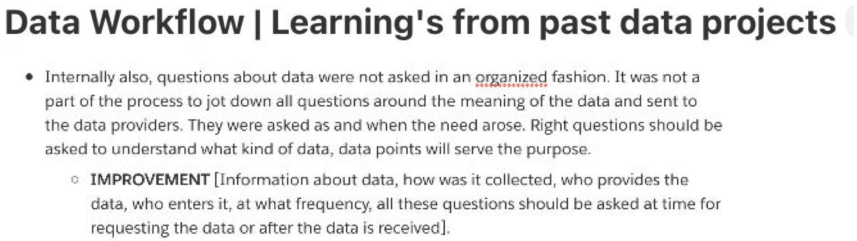
When Mon Sep 18, 2017 1pm – 2pm India Standard Time

Video call https://plus.google.com/hangouts/_/socialcops.com/achyut

Calendar ankita@socialcops.com

Who
- achyut@socialcops.com - organizer
- ankita@socialcops.com
- sourabh@socialcops.com

Going? Yes - Maybe - No more options »



Data Workflow | Learning's from past data projects

- Internally also, questions about data were not asked in an organized fashion. It was not a part of the process to jot down all questions around the meaning of the data and sent to the data providers. They were asked as and when the need arose. Right questions should be asked to understand what kind of data, data points will serve the purpose.
    - **IMPROVEMENT** [Information about data, how was it collected, who provides the data, who enters it, at what frequency, all these questions should be asked at time for requesting the data or after the data is received].

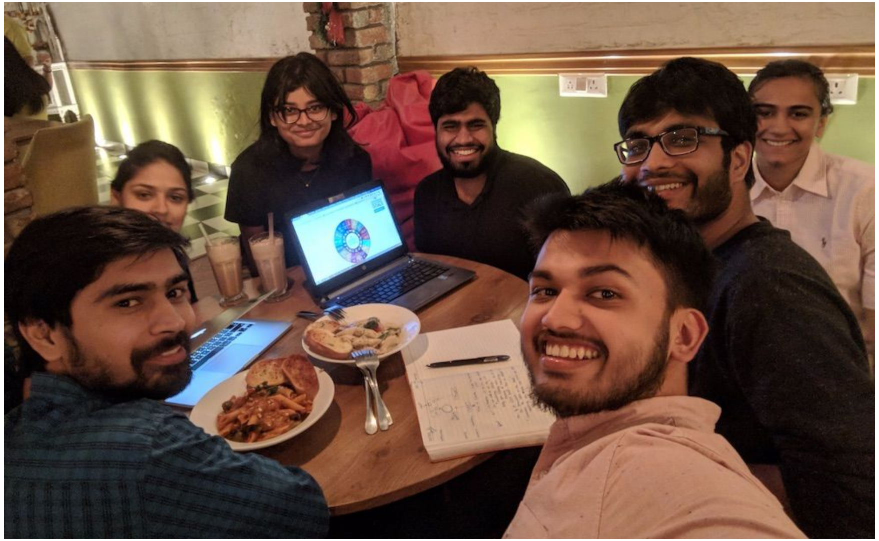
✅ Quarterly Start/ Stop/ Continue Exercises

Building Trust & Collaboration

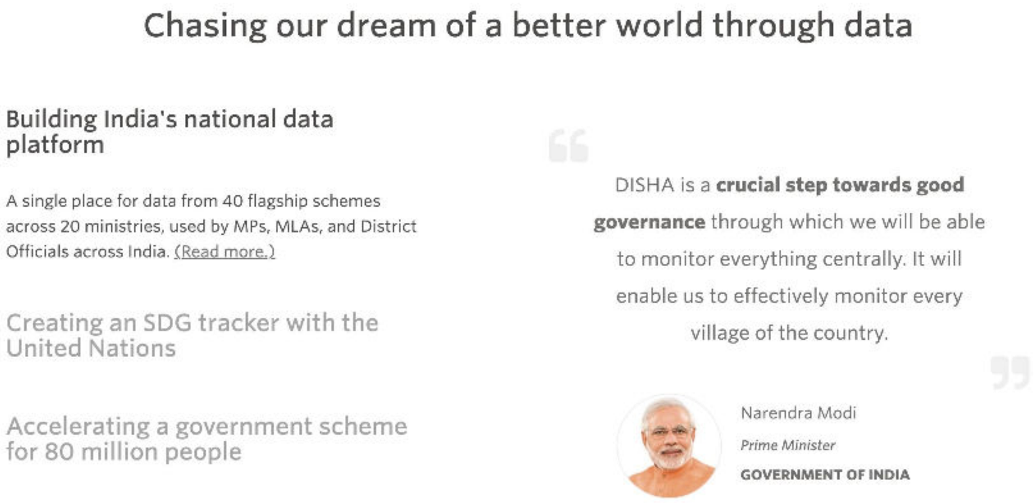
HMW improve collaboration between data engineering & analysts?

HMW to bring problems out in the open & reduce “data frustration

🎉 Data Cribbing Parties



So…. that’s (mostly) the backstory of how our team became 6X more agile in 2 years.



Chasing our dream of a better world through data

Building India's national data platform

A single place for data from 40 flagship schemes across 20 ministries, used by MPs, MLAs, and District Officials across India. (Read more.)

Creating an SDG tracker with the United Nations

Accelerating a government scheme for 80 million people

> DISHA is a **crucial step towards good governance** through which we will be able to monitor everything centrally. It will enable us to effectively monitor every village of the country.

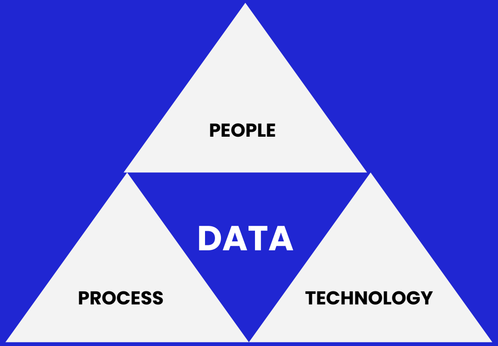
Narendra Modi
Prime Minister
GOVERNMENT OF INDIA

Everything we do, we do because of our experiences as a data team ourselves. Thank you for helping us build Atlan ♥
atlan
Nasdaq
W 43 St
ONE WAY
ONE WAY
Durst

So what does any of this have to do with DataOps?

DataOps is a discipline, not a product

DataOps is emerging as a new discipline— bringing principles of Agile, DevOps and Lean to Data Management



DataOps focuses on binding people, technology and processes to create an agile data culture.

The DataOps Culture Code

🗄️ Treat data, code, models and dashboards as assets

All data assets, from data to dashboards, are assets, and they should be treated like assets.

- Assets should be easily discoverable.
- Assets should be maintained.
- Assets should be easily reusable.

🚀 Optimize for agility

In today’s world, as business needs evolve rapidly, data teams need to be a step ahead, not deluged with three months of backlog.

Constantly measure your team’s velocity, and invest in foundational initiatives to improve cycle times.

- Reduce dependencies between business, analysts and engineers.
- Enable a documentation-first culture.
- Automate whatever is repetitive.

Source: https://docs.atlan.com/our-manifesto/dataops-principles

The DataOps Culture Code

👥 Create systems of trust

With the inherent diversity of data teams, it's all too easy to misunderstand other team members' roles. But that creates trust deficiencies — especially when things go wrong! Intentionally create systems of trust in your team.

- Make everyone’s work accessible and discoverable to break down ‘tool’ silos.
- Create transparency in data pipelines and lineage so everyone can see and troubleshoot issues.
- Set up monitoring and alerting systems to proactively know when things break.

🖇️ Create a plug-and-play data stack

The data ecosystem will rapidly evolve. The tools, technology and infrastructure you use today will (and should) be different from the tools you use two years later.
Your data stack should allow your team to experiment and innovate as technology evolves, without creating lock-ins.

- Embrace tools that are open and extensible.
- Leverage a strong metadata layer to tie diverse tooling together.

Source: https://docs.atlan.com/our-manifesto/dataops-principles

The DataOps Culture Code

✨ User experience defines adoption velocity

Teams at Airbnb famously said, "Designing the interface and user experience of a data tool should not be an afterthought." Without good user experience, the best tools or most thoughtful processes won't be adopted in your team.

Invest in user experience, even for internal tools. It will define adoption velocity!

- Invest in simple and intuitive tools.
- Software shouldn't need training programs.

🤝 It's a team sport — collaboration is key

Data teams will always have a variety of roles, each with their own skills, favorite tools and DNA. Embrace the diversity, and create mechanisms for effective collaboration.

Source: https://docs.atlan.com/our-manifesto/dataops-principles

p@atlan.com

@prukalpa

metadataweekly.substack.com