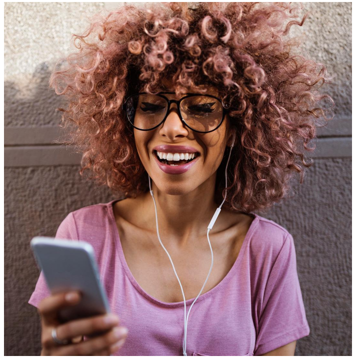
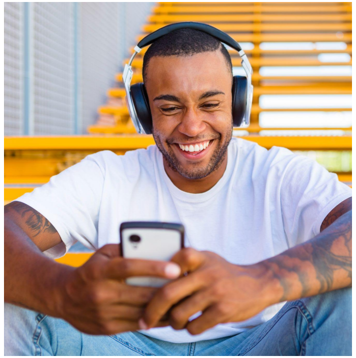
Building large scale analytical products

Shweta Shah
VP, Data Science, Nielsen

Nielsen shapes the world's media and content as a global leader in audience measurement, data and analytics.

Our data underpins essential transactions that propel the buying, selling and creation of media. Through our understanding of people and their behaviors across all channels and platforms, we empower our clients with independent and actionable intelligence so they can connect and engage with their audiences—now and into the future.





Our purpose

Powering

A business built on Audience Is Everything™ delivers impactful media and consumer data and insights to inform today and shape tomorrow.

a better media future

Empowering the world's marketers, creators and innovators with transparent and enlightening audience insights.

for all people

A global company that ensures all voices are heard and helps people create and shape a more open, connected and trustworthy society.

A comprehensive portfolio of solutions

Audience Is Everything™

one measurement solutions

impact marketing solutions

gracenote content solutions

Single media platform

Global, digital-first footprint

Big data validated by real people

Nielsen

Building large scale products

Projects vs products

Project	Product
Fixed goal	Flexible goal
Short term engagement	Long term engagement
Solution for now	Solution for the future
One and done	On-going value-add

Product development challenges

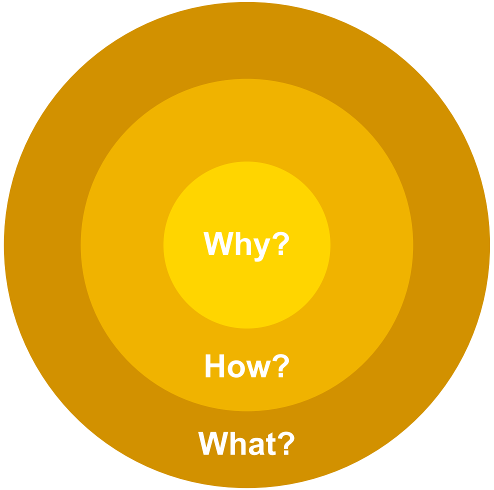
- Large and overwhelming effort
- Unclear and open-ended Goals and requirements
- Flexible and ever-expanding goals
- Longer timelines
- Loss of inspiration

How to continue to keep product development inspiring?

What builds inspiration?

Start with the "why?" - Simon Sinek

The golden circle



Why?

What is your purpose? **Motivation**

How?

How will you achieve your ***why***? **Process**

What?

What is the result of the ***why*** and ***how***? **Product**

Nielsen

Keep the inspiration alive

Bring the golden circle in at all levels of work!

Thin slices - a planning framework

Thin slice your product

How to build your product? - one thin slice at a time

A thin slice is a ***minimal*** but ***robust*** strand of ***end-to-end*** functionality that brings a value-add to the product.

Why thin slice?

1. Breaks a larger goal into smaller and more achievable and inspiring goals
2. Thin slices are end-to-end functionalities - they are not broken down by function
3. Completion of a thin slice aims at producing value-add and revenue realization

How to create a thin slice?

Evaluate the impact of an advertisement on the brand globally

Should we focus on the US only?

Should we focus on a single advertising medium only - like TV?

Should we focus on just National TV outlets rather than all TV outlets?

Evaluate the impact of an advertisement shown on National TV on the brand in the US

Stages of thin slice lifecycle

Idea | Input from stakeholders | Roadmap | Dev teams working on it | Roll out

0
At the start line
First idea

1
Discover and refine
Determine feasibility and exact scope

2
Prioritize
Prioritize on the roadmap

3
Design and implement
Design timeline and tasks
for all development

4
Dev done and UAT testing
Development is complete

Commence User-acceptance testing

5
At the finish line
done - rolled out to
all stakeholders

Nielsen

Discover and refine

Build your golden circle

Need to determine whether the thin slice is a viable idea.

Discovery means:

- Determine the value and viability
- Determine the feasibility

Use a value canvas to determine value, viability and feasibility

Value canvas - value and viability risk

Problem:
Top 3 problems you face

Outcome:
Top 3 outcomes wanted

Unique value prop:
Single, clear, compelling message that states why you are different and worth buying/using

Success:
What tangible criteria defines what success look like?

Assumptions:
Beliefs that have not yet been confirmed related to value, viability, feasibility or usability

Risks:
Implications of constraints/ dependencies

Competitor:
Who are the competitors and what advantage does solving the problem provide?

Customer segments:
Target customers (clients and personas)

Cost structure:
Customer acquisition costs, distribution costs, hosting.

Revenue stream:
Revenue model, lifetime value revenue...

Nielsen

Value canvas - feasibility

Data review?

- *Is real data available?*
- *Is sample available?*
- *Do you have the right data facts and data dimension available?*
- *At what frequency is the data available?*

Profile of data

- *What is the data size and is it sufficient for solution?*
- *What does the data quality look like?*

Tech needs

- *Will the current tech stack support this solution?*
- *Are there architectural changes needed?*
- *Are there any external tech dependencies?*

Tech risks

- *What impact is expected from architectural changes?*
- *Are there any risks to call out?*

Potential risks: tech stack insufficient, resources scale up needed, lack of expertise in technology required, etc

Key data science solutions

- *What are the key ideas of the solutions to evaluate?*
- *For each idea - what is the suitability to answering the question, what are the potential risks, what is the potential cost?*

Data science risks

- *How likely is it that this solution can be implemented/ investigated successfully?*

Potential blockers: lack of information in the data, model is too complex, no expertise available in the technology required, runtime can be too high, etc

Discover and refine

Build your golden circle

If viable, then the idea needs to be refined.

Refine means:

- Align with stakeholders the goal of the thin slice: Value statement
- Make sure the scope is clear: Outcomes/target segments/assumptions/out-of-scope
- Determine what types resources and talents are needed for the development
- Determine how much effort the thin slice would be
- Get values for value, benefit, and impact from the stakeholders.

Design and implement

Define your implementation plan

Create a thin slice design or a project plan

Design means:

- Working out which steps are needed
- Creating a list of solutions to evaluate based on feasibility analysis
- Working out as granular level details as possible on required tasks and stories
- Mapping it all to a timeline

Nielsen

Project plans

Map out your plan to a timeline

Sprint	Start date	End date	Sprint goals
Sprint 1	3/1/2023	3/21/2023	Goal 1: Create mock-up data to test methodologies Goal 2: Create basic architectural design
Sprint 2	3/22/2023	4/11/2023	Goal 1: Make a decision on which methodology to use Goal 2: Create input data pipeline
Sprint 3	4/12/2023	5/2/2023	Goal 1: Imprement <based on methodology selected> Goal 2: Test <based on methodology selected>
Sprint 4	5/3/2023	5/24/2023	Goal 1: Design output structure and pipeline

Purposeful data science - An implementation framework

What is “purposeful” data science?

Put **the purpose** at the heart of how we work - bringing the golden circle into everything we do

(meaning) When we communicate we should put the purpose, or the ‘why’, on equal footing with the ‘what’.

(intention) When starting a task we should reflect on what is meant to be produced and partition our work with that mind.

Purposeful data science is an implementation framework for teams working with data scientists.

Nielsen

Purposeful Data Science Framework

Effective communication

- Lack of clear communication is a challenge for cross functional teams
  - It is not efficient to communicate all details to all disciplines, or even all team members within a discipline.
  - Work cannot be effectively prioritized without understanding.
- The purpose of “why” you are working on something is more universally understandable than the specifics of “what” you are working on.

Update every task with

- A “why” - the purpose behind the task
- A “how” - how the is going to be done task
- A “what” - the final deliverable of the task

Purposeful Data Science Framework

Efficient work breakdown

- Data science work, specially research, is often open ended without a clear definition of done
- Research, is a series of decisions that leads to a final solution
- Thinking about your work as a series of tasks to produce these decisions helps create an efficient breakout of work

Categorize your tasks based on what you are trying to produce

- Information, in terms of an answer to a specific question
- Code, in terms of a standalone module
- (Model) artifact, in terms of a serialized model object

Let’s work through an example

Bucketing consumers based on purchase behaviour

Old approach - task focused

1. Perform EDA on the data
2. Identify clustering algorithm
3. Implement in code

New approach - purpose focused

1. Information
   a. Which columns represent the purchase behaviour?
   b. Is there any data missing and how to deal with it?
   c. What is the list of clustering algos I want to try?
   d. What are the edge cases I need to consider?

2. Code
   a. Code to implement clustering algo chosen
   b. Code to account for edge cases identified
   c. Integration into bigger pipeline

3. Artifact
   b. Create and analyse consumer buckets for specific test cases

Summary

Thin slice

Achievable inspiring goals

End-to-end strand of work that add value

Clear scope definitions with well-defined timelines

Purposeful data science

Helps with efficient communication

Provides a way to break out work more clearly

Converts open-ended research questions to purposeful tasks

Golden circles are at the heart of inspirational development

Nielsen

Appendix

Data science is a long series of decisions

Partitioning work: Producing information

Be explicit about the question that you are answering.

Definition of this task type:

- Working on providing an answer to a specific question by collecting insights that will help arrive at a recommendation and decision. Insights can include finding and exploring a coding package, reading academic research, asking other stakeholders, exploring the data to understand what types of further analyses are needed, or building exploratory models.

When can this task be considered complete?

- The question originally asked has an answer and the answer is readily available and provenance of that answer is known. This documentation should include assumptions and key decisions made to get to the answer.
- The documented answer has been reviewed and accepted by the Data science lead.

Without code you don’t have a product

Partitioning work: Producing code

Our products only exist if they are implemented in code and in a version control system.

Definition of this task type:

- Write documented/commented code that everyone in the team can run and obtain the same output.

When can this task be considered complete?

- Code has been reviewed and merged
  - Alignment of expectations of code across teams is an open discussion but as a starting definition:

    The code runs and produced the same result for both the author and the reviewer(s), the contribution is understandable by a reviewer not entirely familiar with the codebase (i.e. it is well documented), the integration with the larger project is successful (e.g no other code breaks due to the addition/changes) and, if it doesn't, new issues are generated to account for that. The measure of how reproducible the result is can also come in form of tests being passed. If the goal was some sort of performance enhancement (time, memory, accuracy), those criteria have to be met.

It’s easy to just hit “run”

Partitioning work: Producing model artifacts

Product development also requires tasks that don’t produce information or produce code, but need the code to run to test the solution, reproduce a problem, or simply create a result or dataset.

Definition of this task type:

- Setting the configurations and running the entire process that produces a metric or some other output required for the product, required to test a solution, or required to reproduce a problem.

When can this task be considered complete?

- Artifact of interest is produced along with documentation