



# Cooking up a new search system

Recipe search at Cookpad

**Berlin Buzzwords**

19 June 2023

**Matt Williams**

Principal Engineer – Search and Discovery  
at **Cookpad Global**

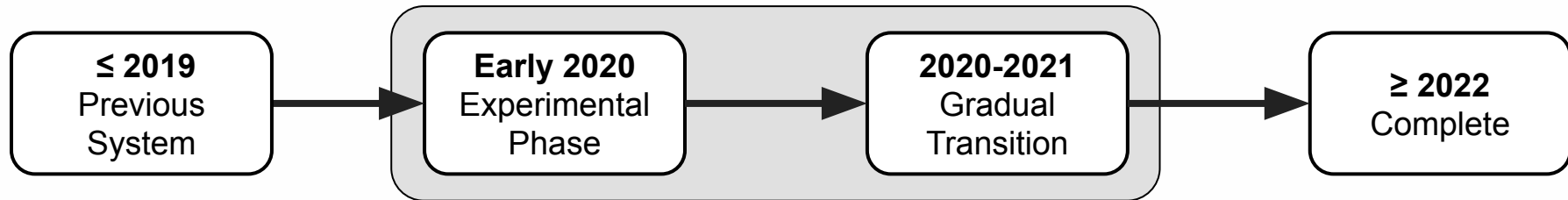
**Web:** [mattjw.net](http://mattjw.net) | **Social:** @voxmjw | **Cookpad:** @matt

**"A search team should be responsible for optimising search relevance and improving the search experience."**

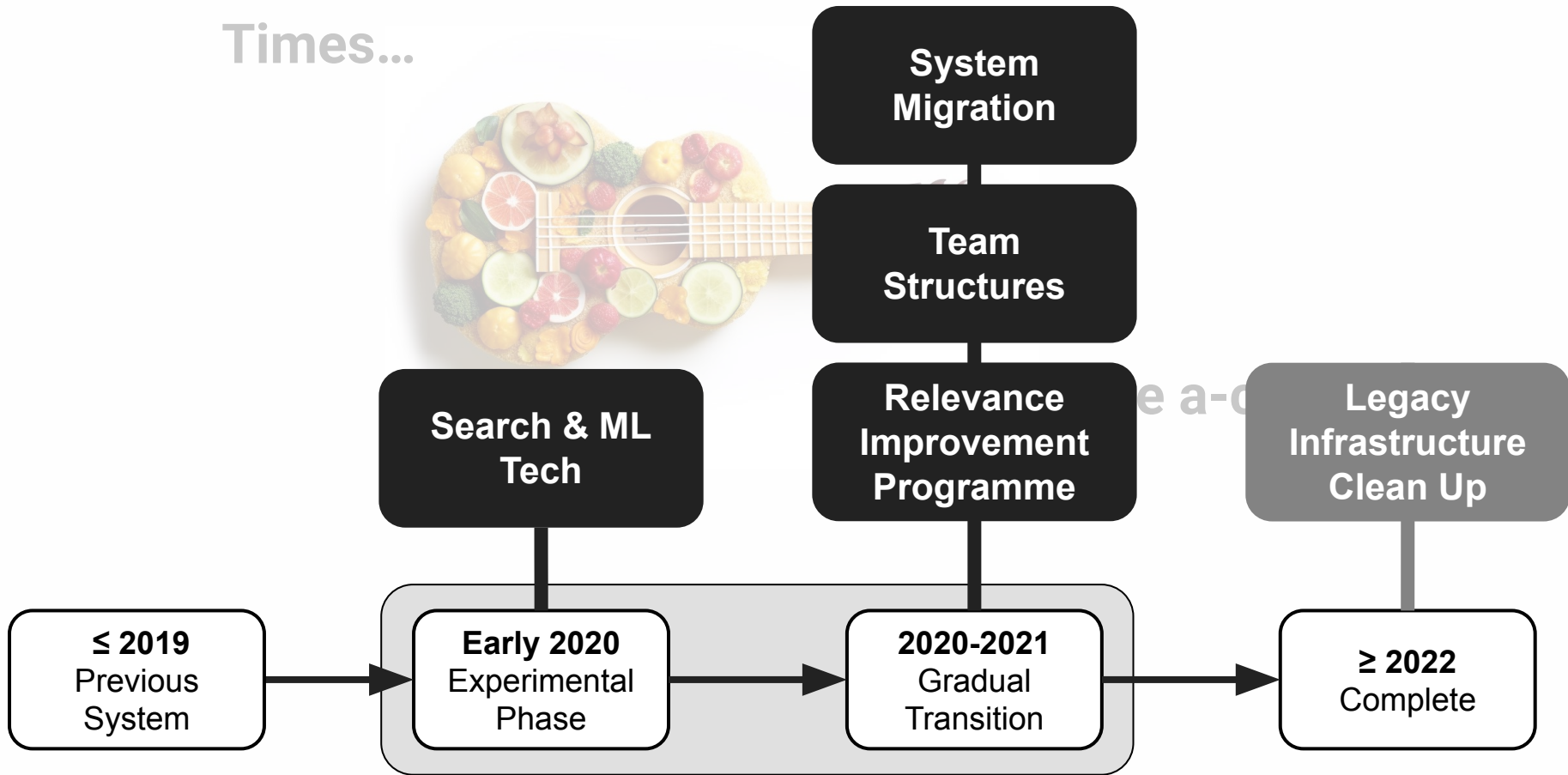
# Times...



## ...they were a-changin'



Times...



Your Network
Inspiration

### What's in your fridge?

Pick up to 2 ingredients

Rice

bacon

Chickpeas

Potato

Beetroot

Pasta

Spinach

Beef

All
chickpea cutlet
chickpea salad
chickpea falafel

**Beetroot Hummus** ❤️

4 9 1

Save

### Browse Cookbooks

←
beetroot hummus
×

Latest
Popular

### Beetroot and mint hummus

Kate Carless

raw beetroots, can chickpeas, garlic, stems of mint leaves, ground cumin, Zest and juice...

### Beetroot Humous

Lisa Clarke

cooked chickpeas, garlic clove, olive oil, tahini, Salt and pepper, small cooked be...

### Verity of Falafels with beetroot hummus #zoo...

Binaka panchmatia

chick peas (drain water), sprouted green grams, rice flour, semolina, corn flo...

←

## Beetroot and mint hummus

**Kate Carless**  
@KateCarless  
Bath, UK

This got our dinner guests very excited at the weekend because of its incredibly beautiful colour! A great way to use raw beetroots, which when roaste...

MORE

⌚ 40 minutes (includes roasting time for beetroot)

## Cookpad Global...



**50M+ monthly users**



**6M+ recipes**



**70+ countries**



**30+ languages**

**Countries with recently  
published recipes**

## Search...



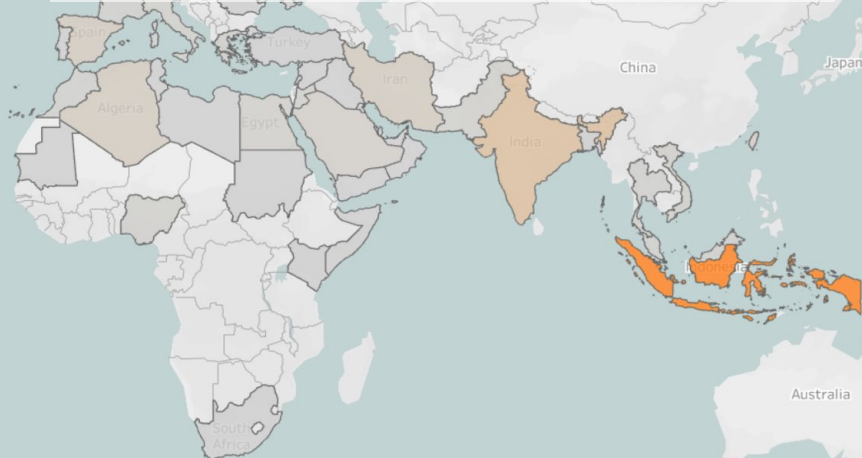
**1M+ searches per day**



**200k+ unique queries per day**



**20M+ requests per day**



**Note: Does not include Cookpad Japan,  
which is a separate app.**

# The end goal

"Search teams and search engineers will **identify impactful opportunities** to optimise search, **build solutions** to solve those problems, and **rigorously validate** the impact."

And a few key ingredients 🍍 🍏 🍇 that helped us get there...

**Responsibility and ownership** over the search experience  
(including **relevance**)



**Autonomy and empowerment** to build solutions  
(while managing cognitive load)

# The end goal

"Search teams and search engineers will **identify impactful opportunities** to optimise search, **build solutions** to solve those problems, and **rigorously validate** the impact."

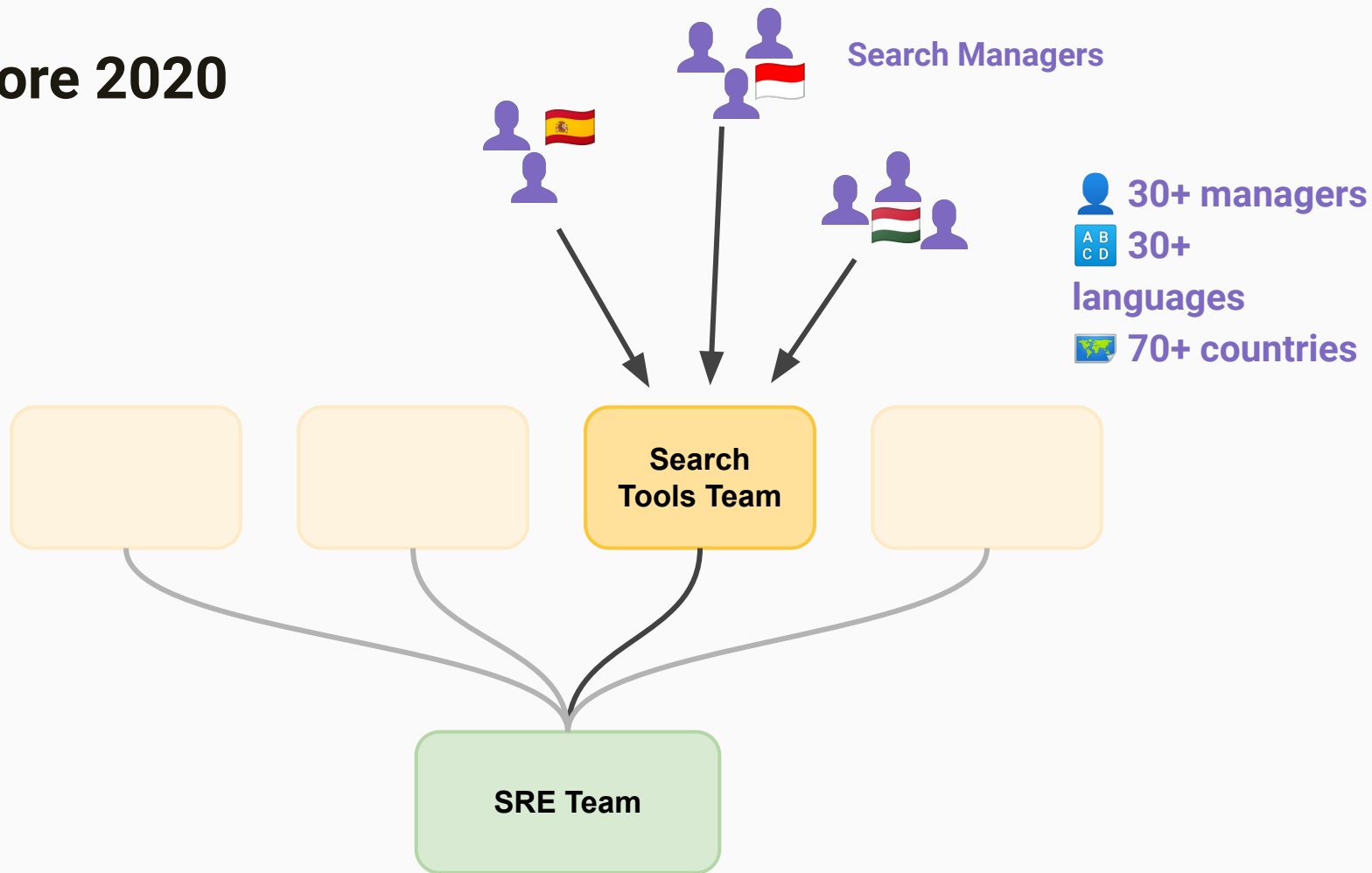
And a few key ingredients 🍍 🍏 🍇 that helped us get there...

**Responsibility and ownership** over the search experience  
(including **relevance**)

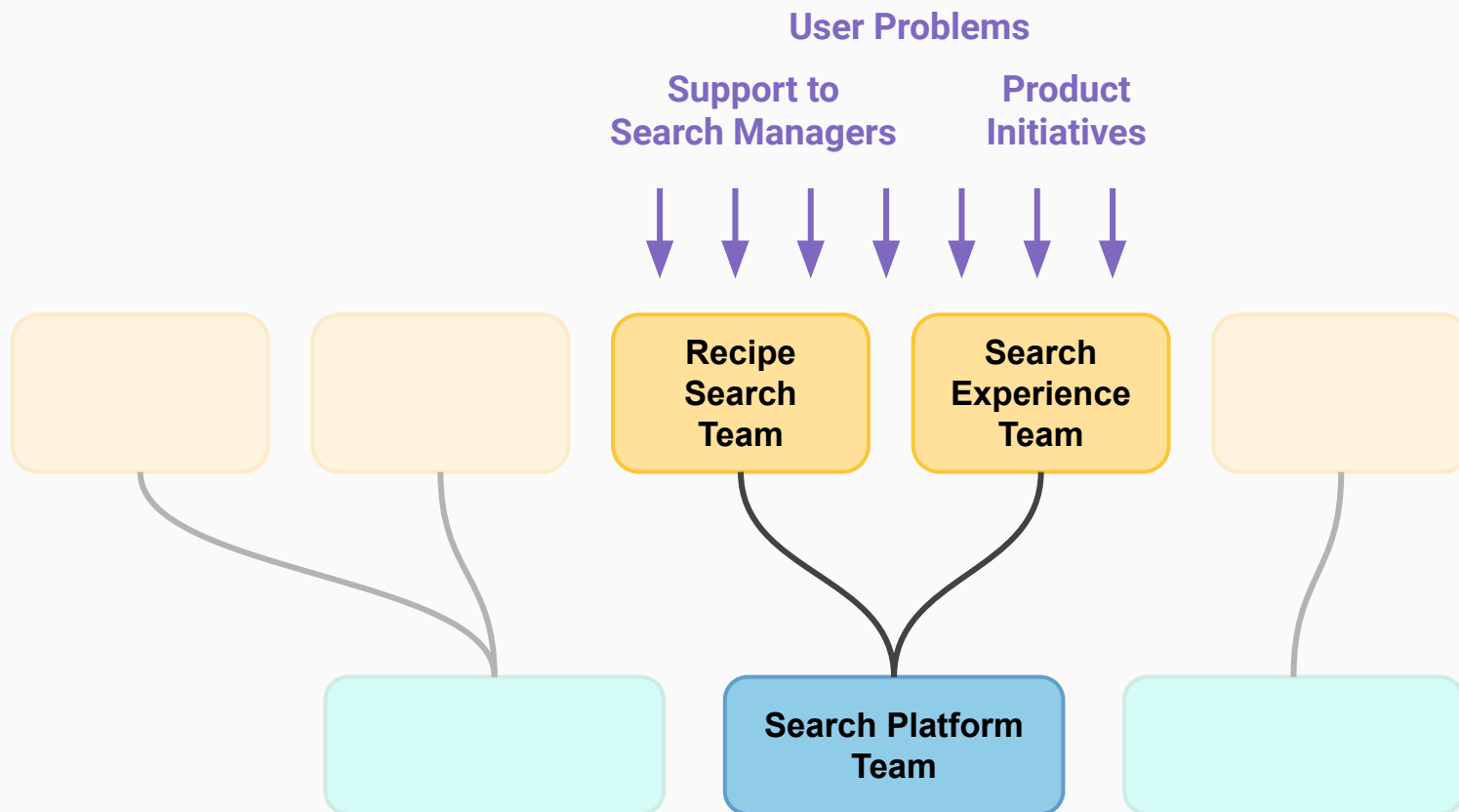


**Autonomy and empowerment** to build solutions  
(while managing cognitive load)

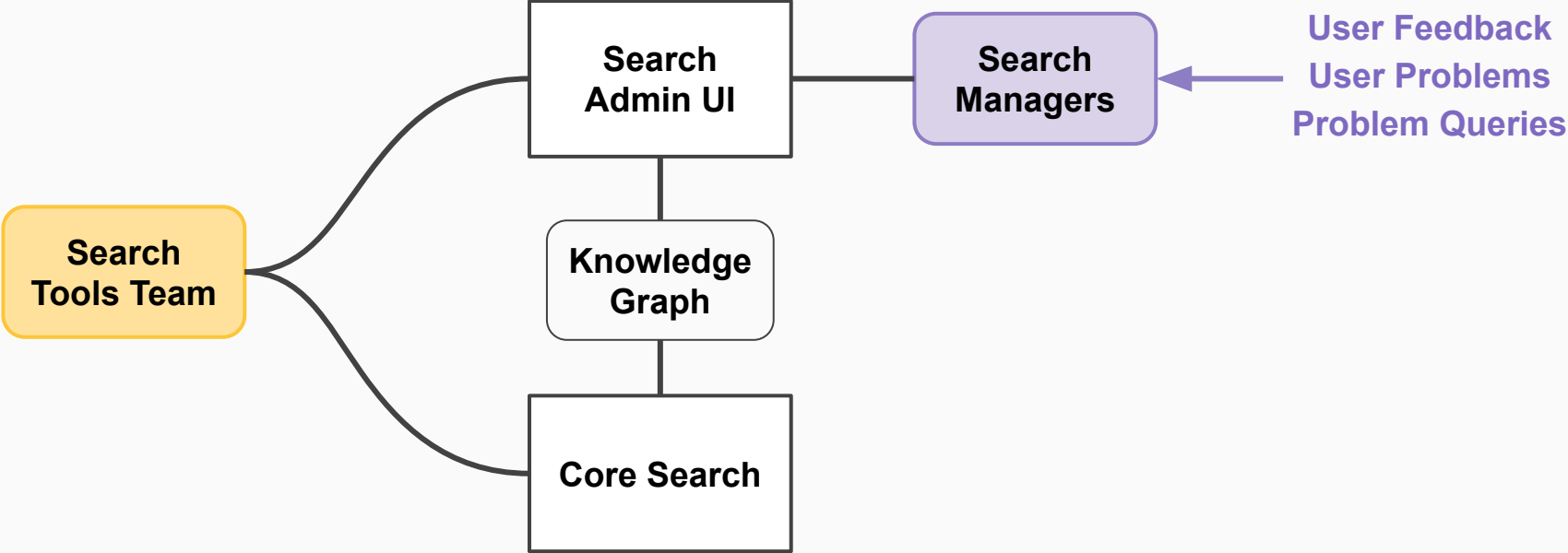
# Before 2020



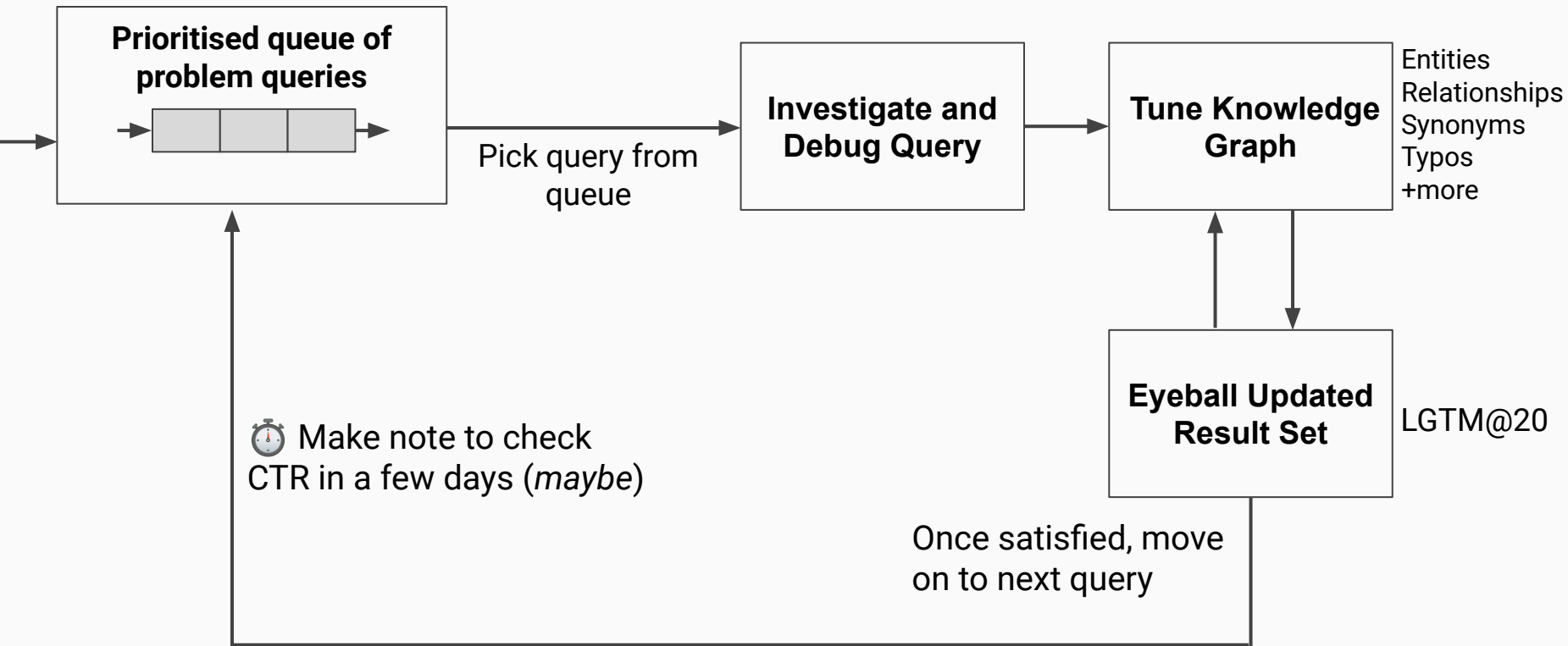
# Target team structure (achieved in 2021)



# Before 2020: Responsibility for relevance optimisation



# Search manager workflow

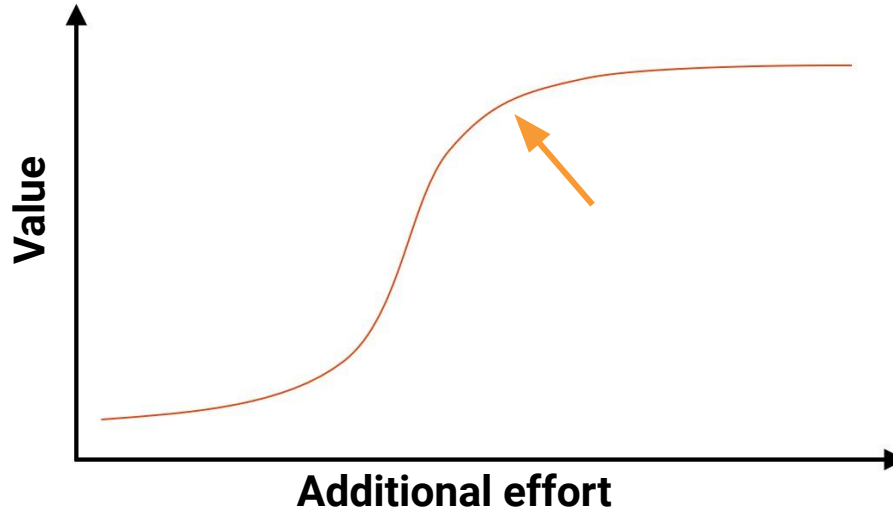


# Limitations



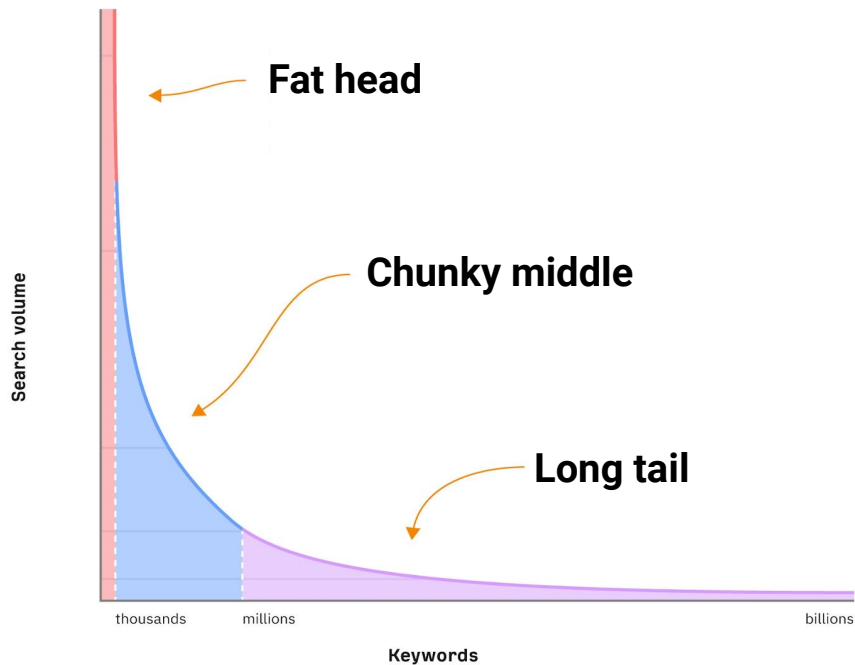
**Consecutive tests, rather than  
controlled experiments**

# Limitations



**Diminishing returns of  
knowledge graph optimisation**

# Limitations

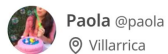


**Trapped in the head of the query distribution**

# The long tail of recipe search queries



Galletas de dinosaurios 🦖



Paola @paola  
Villarrica

<https://cookpad.com/es/recetas/14591385-galletas-de-dinosaurios>  
by Paola

**Head queries**

**≤ 10% of all searches**



pollo

[chicken]



nasi goreng

[fried rice dish]

**Tail (maybe some torso, too...)**

**≥ 50% of all searches**

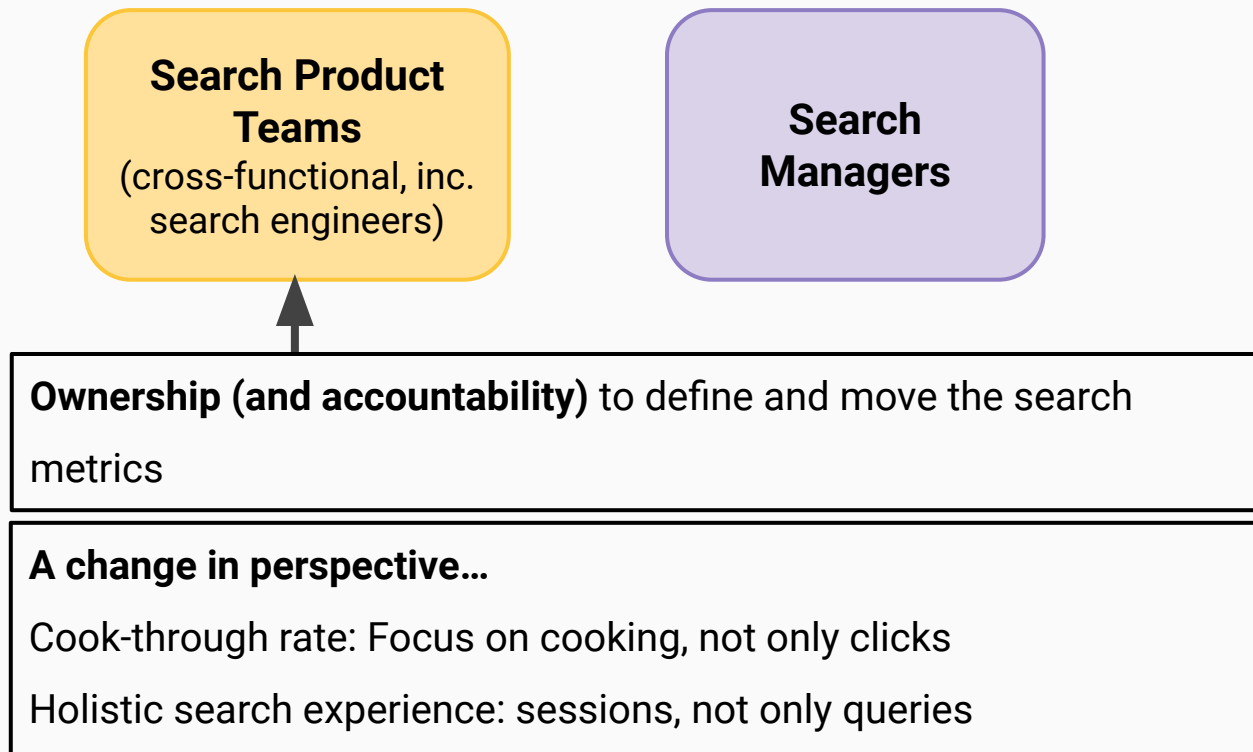


nasigoreng



gourmet cherry blossom  
rice-balls for bento

# Moving the responsibility



# The best of both

Own the models, methods,  
and metrics

Explore data and  
hypotheses at scale  
(size the opportunity)

Controlled experiments at  
scale

## Search Product Teams

(cross-functional, inc.  
search engineers)



## Search Managers

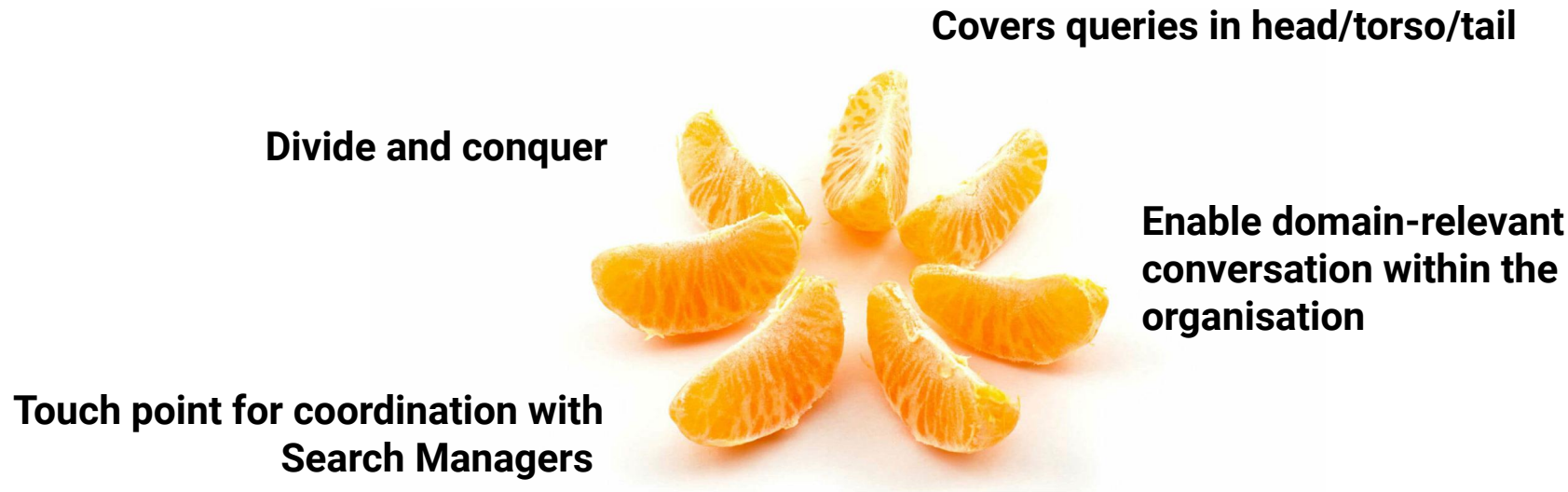
Deep knowledge of local  
culture and cooking

Fluent in local languages

Frequent contact with  
users

How do we get to the **best of both**, while **working in tandem**?

# Query segments: Group queries into domain-relevant cohorts



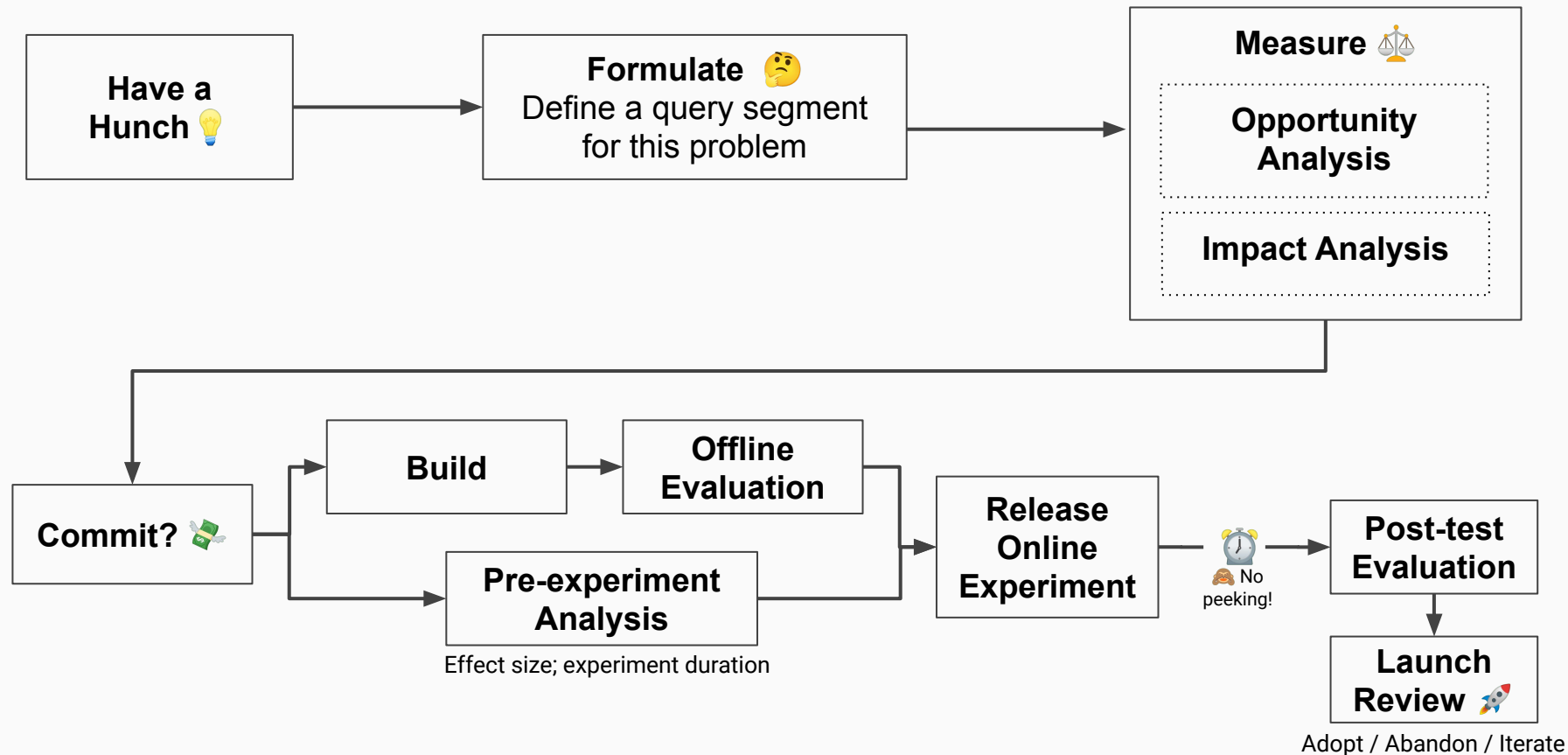
**Example segment:** Single ingredient queries. 🍊 🌿 🌶️

# Segment experiment description card

(example from pre-launch phase of a 2021 experiment)

<b>Name</b>	Single ingredient queries.	<b>Intent description</b>	Searcher wants to cook with this ingredient.
<b>Classification method</b>	Ingredient entities from NER	<b>Regions</b>	Five countries Three languages
<b>Opportunity and impact analysis</b>			
<b>Opportunity size</b>	<b>Searches:</b> $\geq 15\%$  And also: <b>num. queries; num users.</b>	<b>Current quality</b>	<b>Click through rate:</b> <b>4%</b> below avg. <b>Low hits rate:</b> <b>4%</b> above avg.  And also: <b>mAP, session conversion, <u>cook-through</u> rate.</b>
<b>Potential uplift (hypothesised)</b>	<b><math>\geq 2\%</math></b> uplift on CTR (target uplift, before running online experiment)		

# Idealised new search relevance workflow ( $\geq 2021$ )



## Hunch

An idea of an area that  
might be a problem for  
the product

## Where do the hunches come from?!



"Search engineering is not product- and  
user-centric enough. It's all about  
metrics."

A criticism I occasionally hear from  
newcomers to search

**Segments**

**Metrics**

**Opportunity size**

**Impact analysis**

**Experiment design**

**Models and techniques**




**Weekly query triage**


**User interviews**

**Check-ins with Search Managers**

**Dogfooding** (be your user!)

**In-app qualitative feedback form**

 Run online experiments often. Target:  
Evaluate 1+ experiment per week.

 All engineers rooted in real  
problems. See real queries at least  
once per week.

**The language user interface:** Every single query is a unique expression of a human need!



stik keju renyah empuk gurih  
sederhana anti gagal rasa keju banget



easiest method to make a cheese stick, is proven (anti fail),  
and has the dominant taste of cheese

<b>Dish</b>	Specific name of dish	"cheese stick"
<b>Difficulty</b>	Queries which have a difficulty level	"easiest method" / "anti fail"
<b>Taste</b>	Query specifying the taste of the dish	"dominant taste of cheese"

# The end goal

"Search teams and search engineers will **identify impactful opportunities** to optimise search, **build solutions** to solve those problems, and **rigorously validate** the impact."

And a few key ingredients 🍍 🍏 🍇 that helped us get there...

**Responsibility and ownership** over the search experience  
(including **relevance**)



**Autonomy and empowerment** to build solutions  
(while managing cognitive load)

# The end goal

"Search teams and search engineers will **identify impactful opportunities** to optimise search, **build solutions** to solve those problems, and **rigorously validate** the impact."

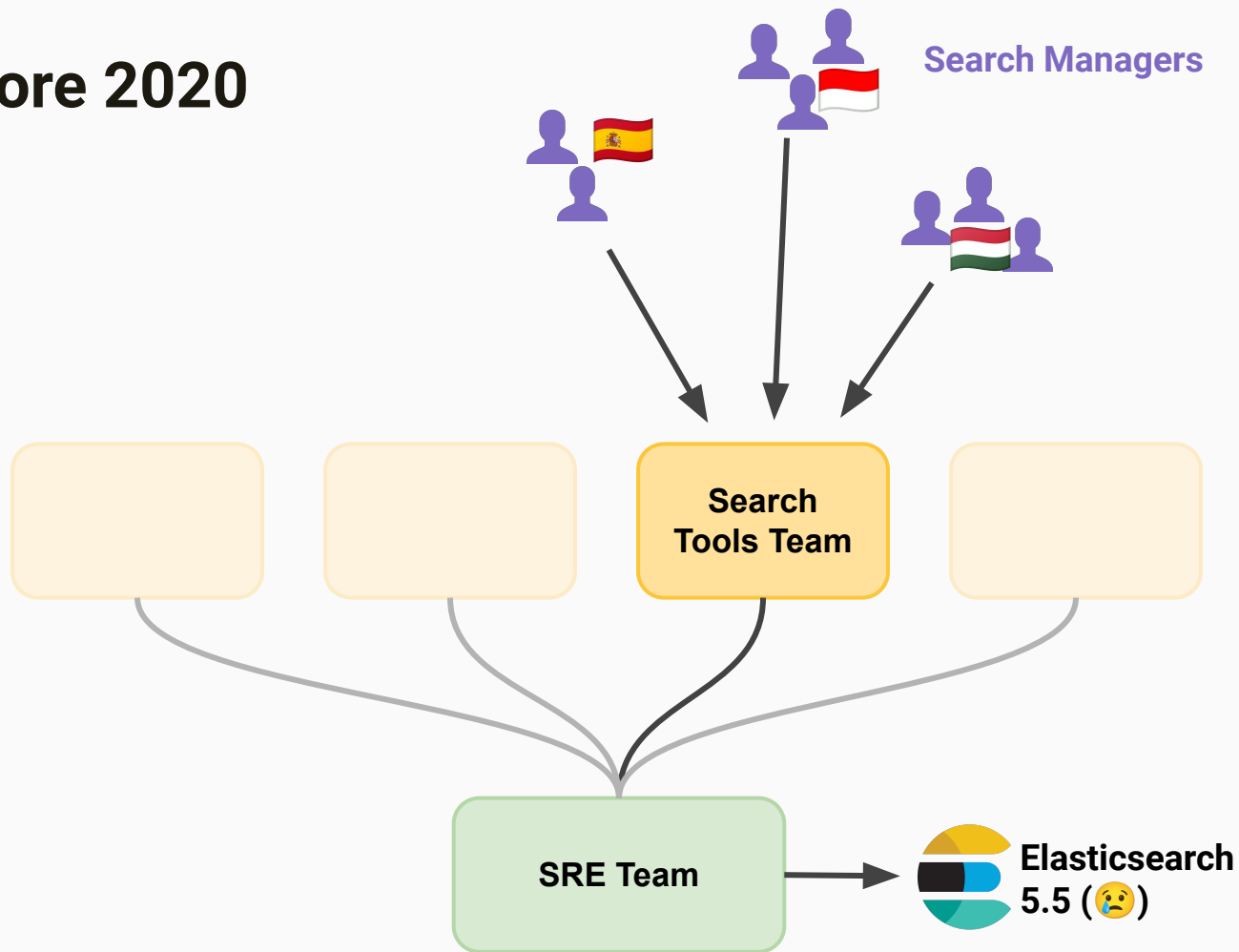
And a few key ingredients 🍍 🍏 🍇 that helped us get there...

**Responsibility and ownership** over the search experience  
(including **relevance**)



**Autonomy and empowerment** to build solutions  
(while managing cognitive load)

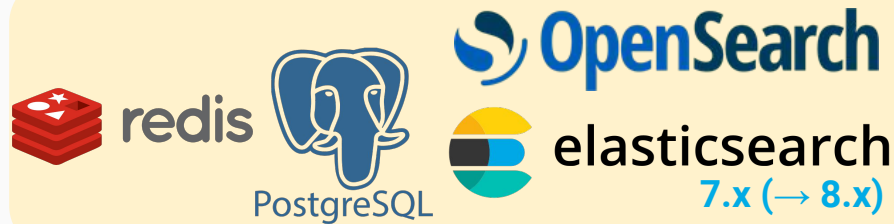
# Before 2020



## Search stack before 2020

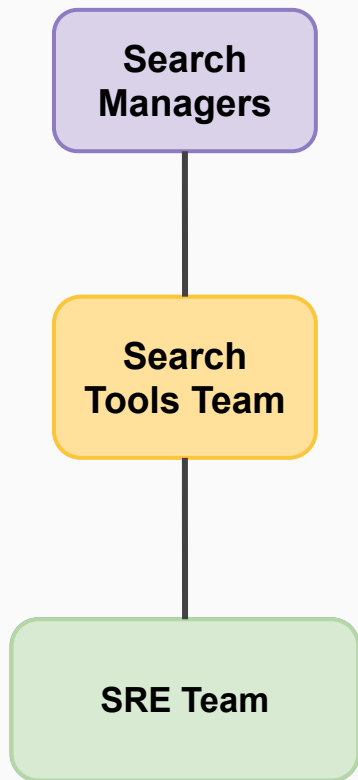


## 2022 onwards

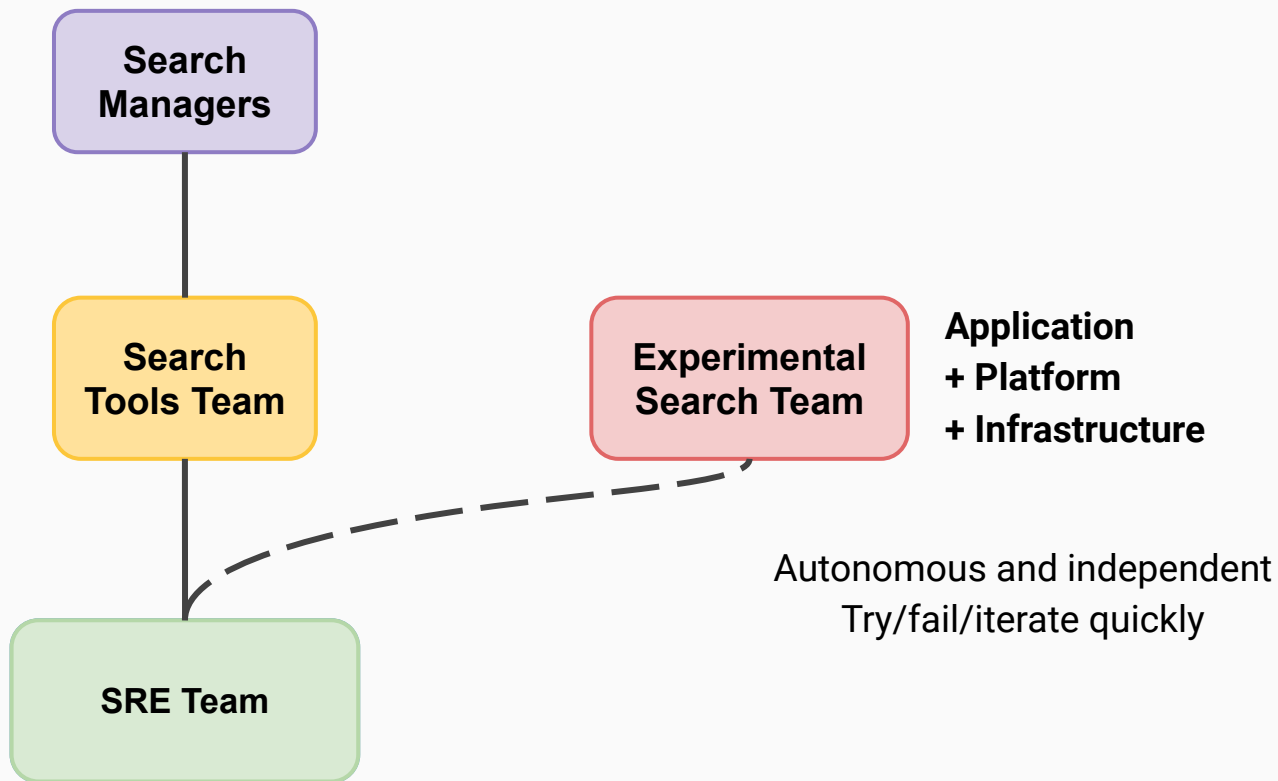


Not pictured: Outside of the search system there is, Ruby, Rails, iOS, Android, and more

# Initiating the change (2020)



# Initiating the change (2020)



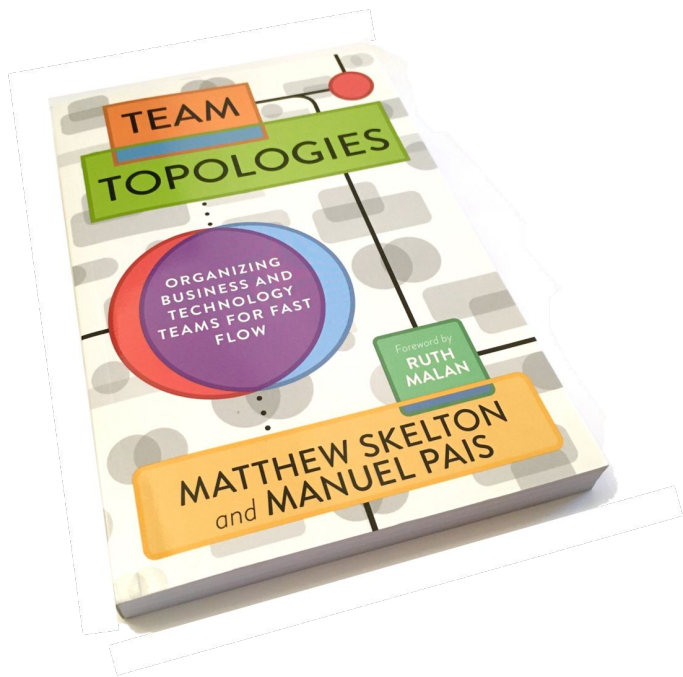
# Initiating the change (2020)



```
Caused by: java.lang.IllegalArgumentException: The value ([Strong bread flour,0,250 grams]) of the type
(org.elasticsearch.spark.sqlScalaEsRow) cannot be converted to an array of
struct<name#text_short:string,position#int:int,quantity#text_short:string>
  at org.apache.spark.sql.catalyst.CatalystTypeConverters$ArrayConverter.toCatalystImpl(CatalystTypeConverters.scala:173 undefined)
  at org.apache.spark.sql.catalyst.CatalystTypeConverters$ArrayConverter.toCatalystImpl(CatalystTypeConverters.scala:154 undefined)
  at org.apache.spark.sql.catalyst.CatalystTypeConverters$CatalystTypeConverter.toCatalyst(CatalystTypeConverters.scala:103
undefined)
  at
org.apache.spark.sql.catalyst.CatalystTypeConverters$ArrayConverter$anonfun$toCatalystImpl$2.apply(CatalystTypeConverters.scala:164
undefined)
  at scala.collection.TraversableLike$anonfun$map$1.apply(TraversableLike.scala:234 undefined)
  at scala.collection.TraversableLike$anonfun$map$1.apply(TraversableLike.scala:234 undefined)
  at scala.collection.Iterator$class.foreach(Iterator.scala:891 undefined)
  at scala.collection.AbstractIterator.foreach(Iterator.scala:1334 undefined)
  at scala.collection.IterableLike$class.foreach(IterableLike.scala:72 undefined)
  at scala.collection.AbstractIterable.foreach(Iterable.scala:54 undefined)
  at scala.collection.TraversableLike$class.map(TraversableLike.scala:234 undefined)
  at scala.collection.AbstractTraversable.map(Traversable.scala:104 undefined)
  at org.apache.spark.sql.catalyst.CatalystTypeConverters$ArrayConverter.toCatalystImpl(CatalystTypeConverters.scala:164 undefined)
  at org.apache.spark.sql.catalyst.CatalystTypeConverters$ArrayConverter.toCatalystImpl(CatalystTypeConverters.scala:154 undefined)
  at org.apache.spark.sql.catalyst.CatalystTypeConverters$CatalystTypeConverter.toCatalyst(CatalystTypeConverters.scala:103
undefined)
```

**SRE team**





## Team Topologies

Matthew Skelton and Manuel Pais

September 2019

And also "**What I Talk About When I Talk About Platforms**"

Evan Bottcher. 2018

<https://martinfowler.com/articles/talk-about-platforms.html>

## User Problems



Search  
Team 1

Search  
Team 2

Internal Development  
Platform

Search Platform  
Team

Maximise their focus on user problems (user value stream)

Own product-relevant **service level objectives**

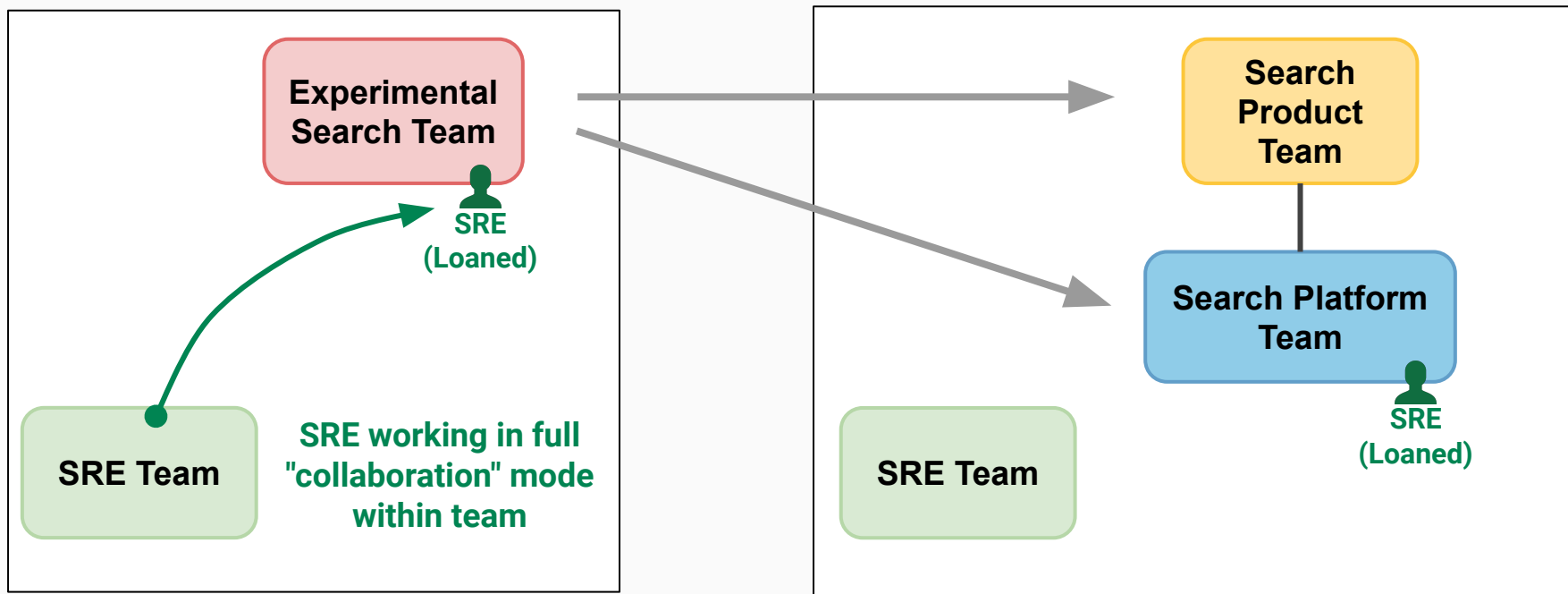
**Autonomy** to introduce new technologies and solutions (and **responsibility** for them)

**"You build it, you run it, you own it"** with help from the platform

Reduce product teams' **extraneous cognitive load**, while enabling solution ownership, through an **IDP**

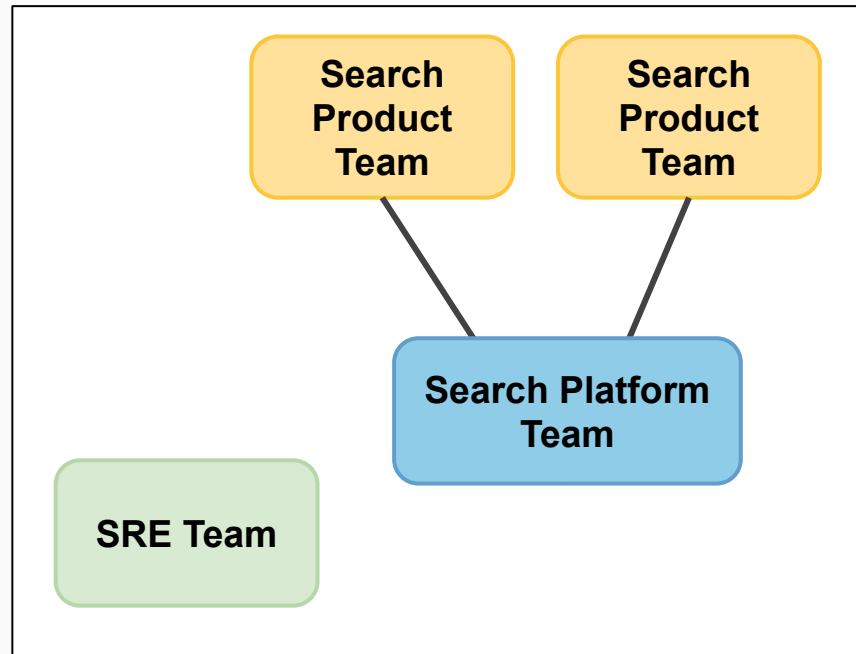
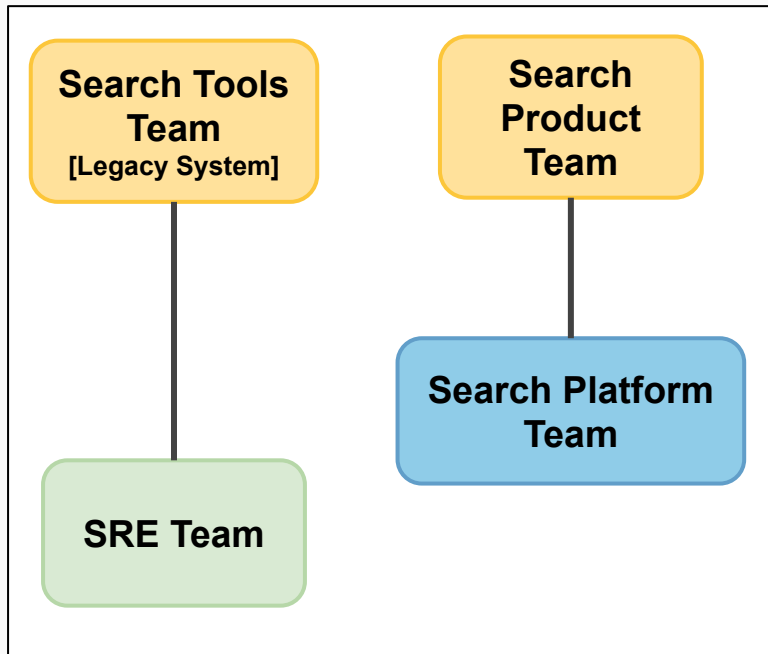
**"As a service"** offerings to product teams

Provide tooling, automation, APIs, and more



## Transition (6 months)

Discover and define relationship between Platform and Product. Establish support practices.



**Final Team Transition**

## User Problems

Support to  
Search Managers

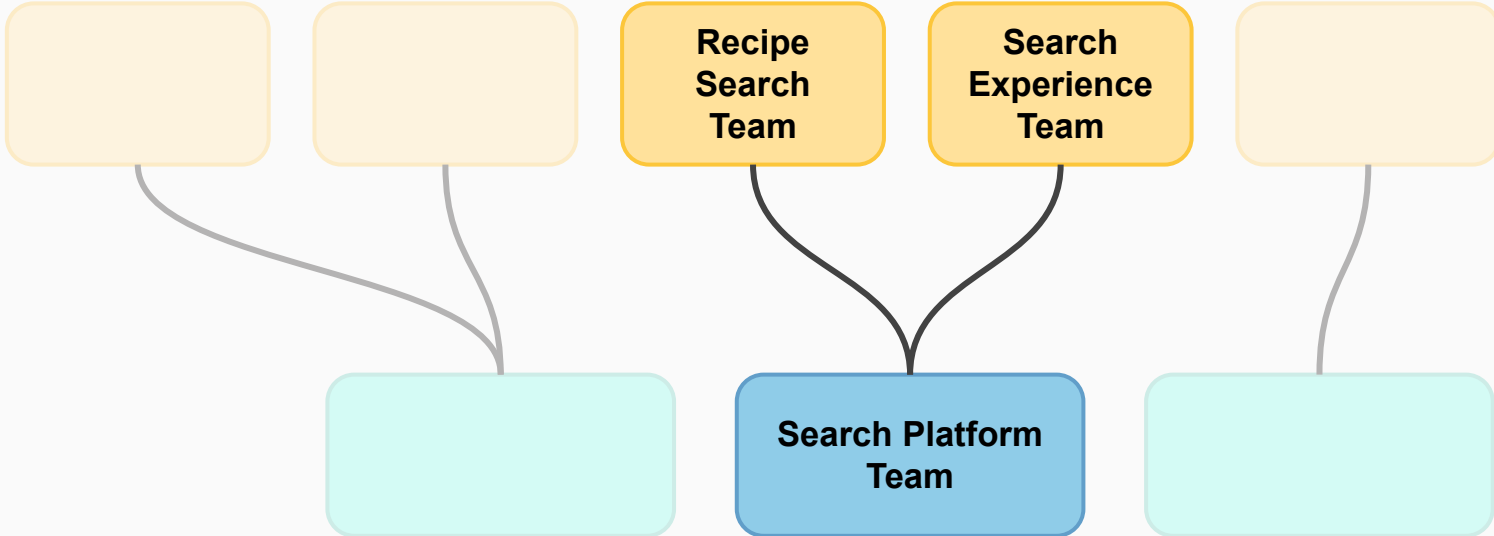
Product  
Initiatives



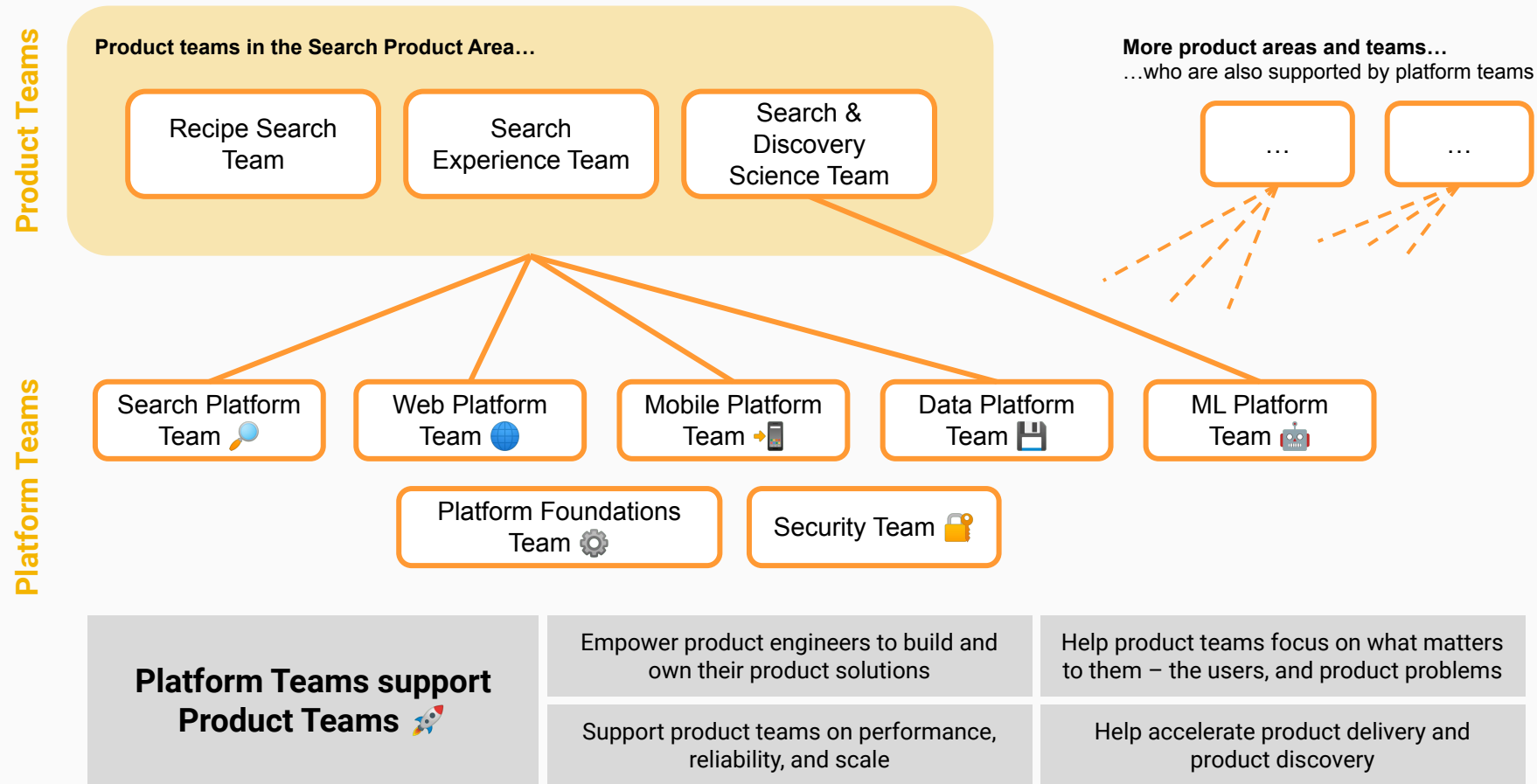
**Recipe  
Search  
Team**

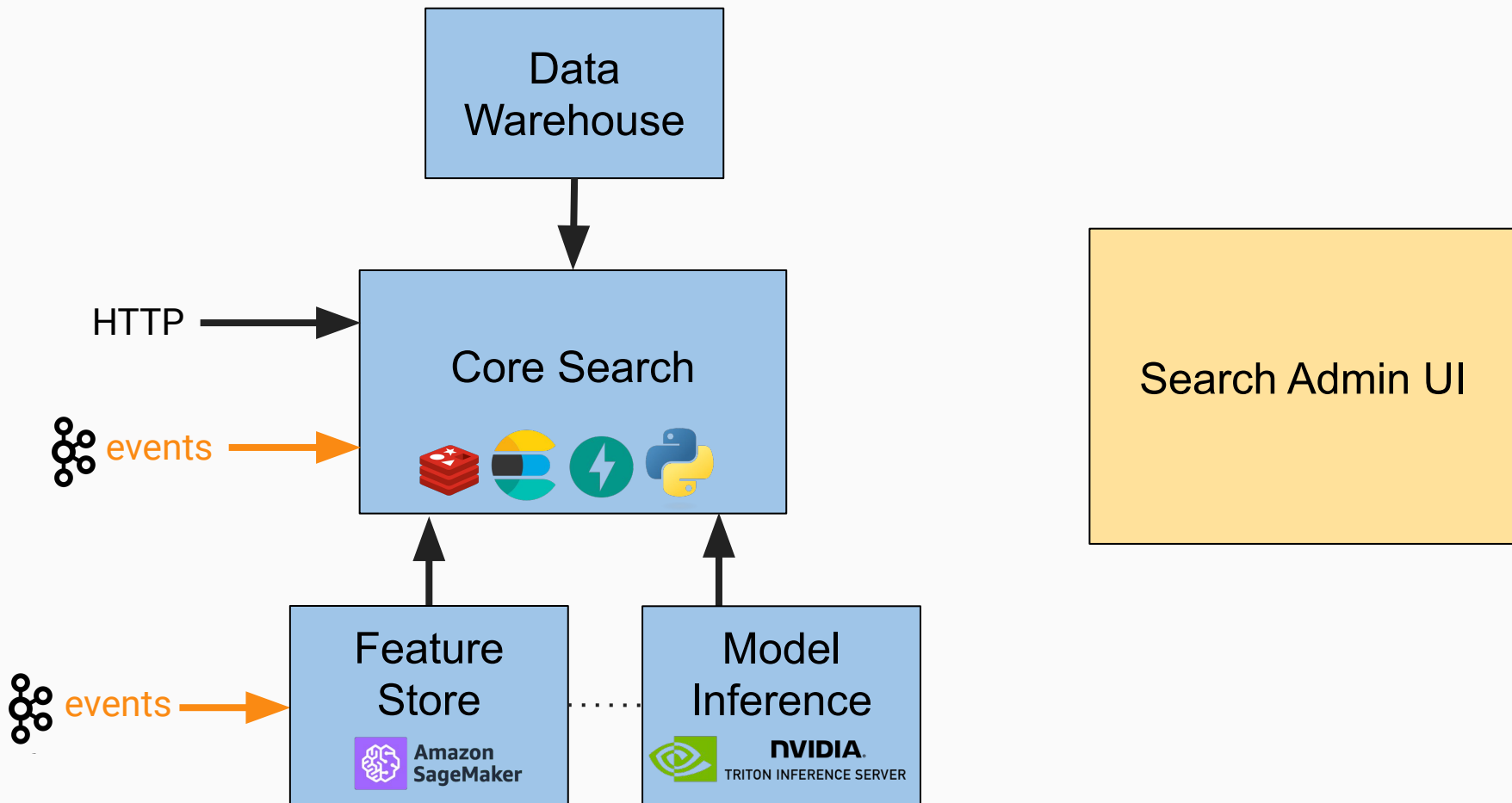
**Search  
Experience  
Team**

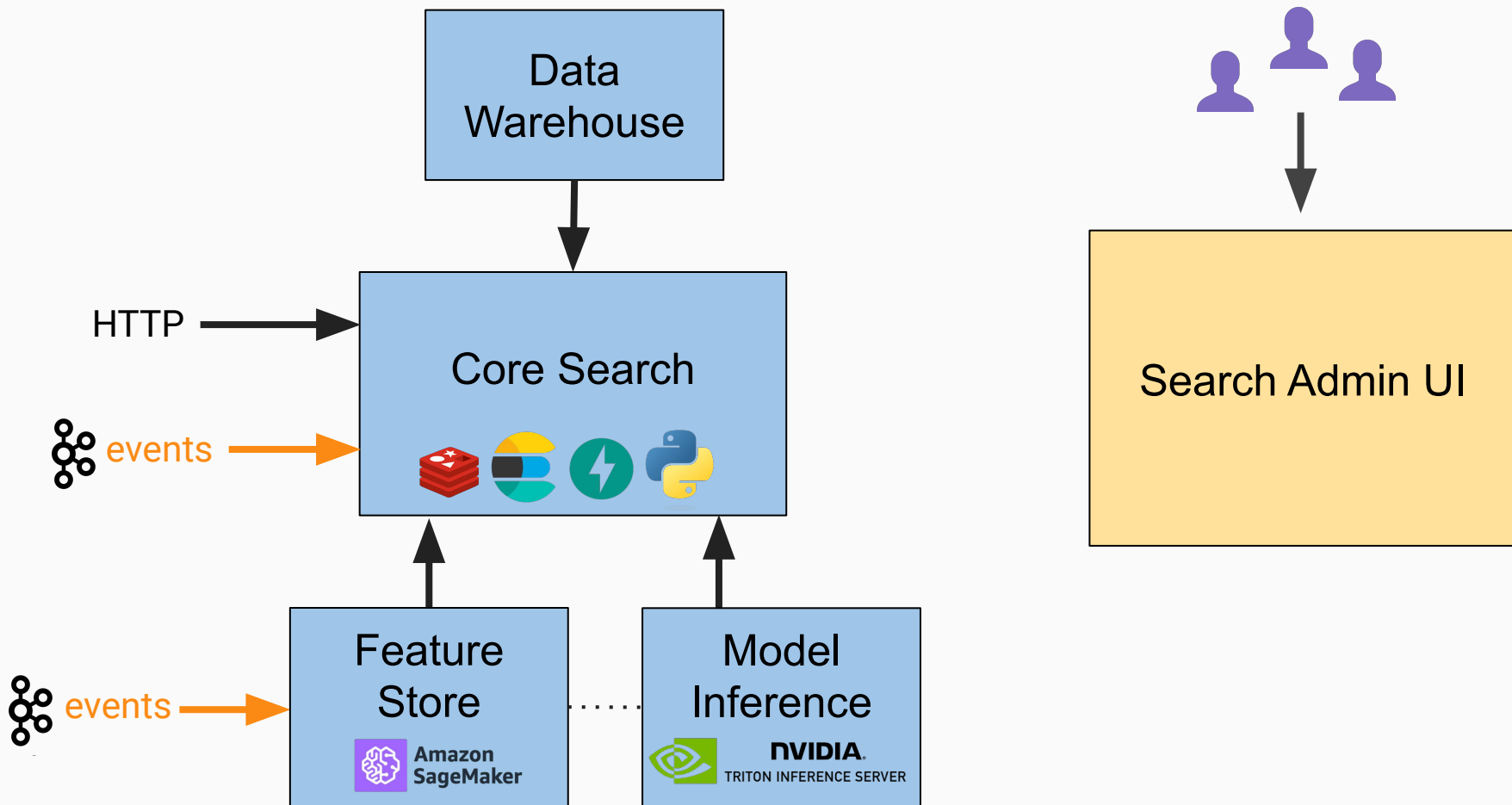
**Search Platform  
Team**

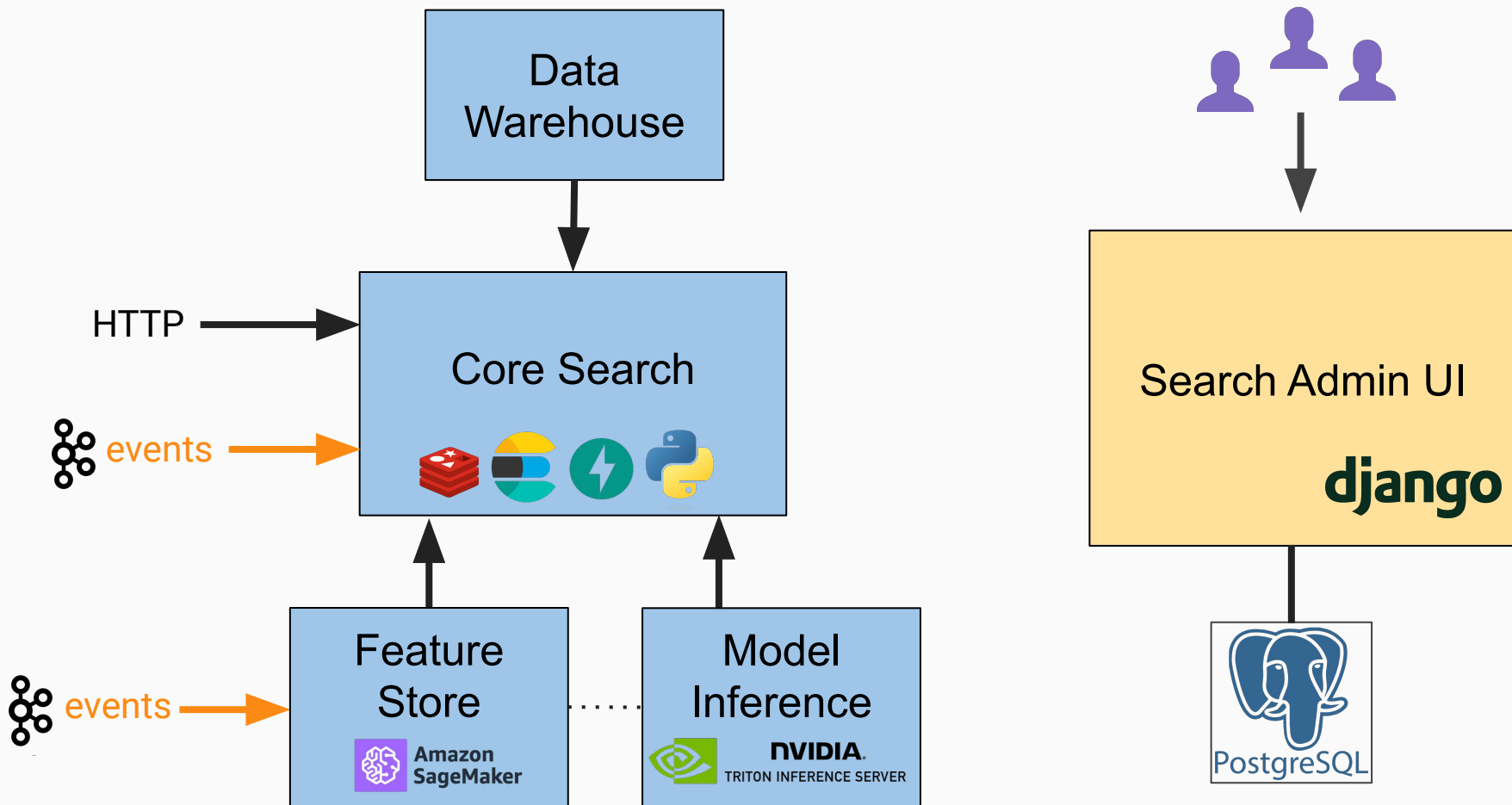


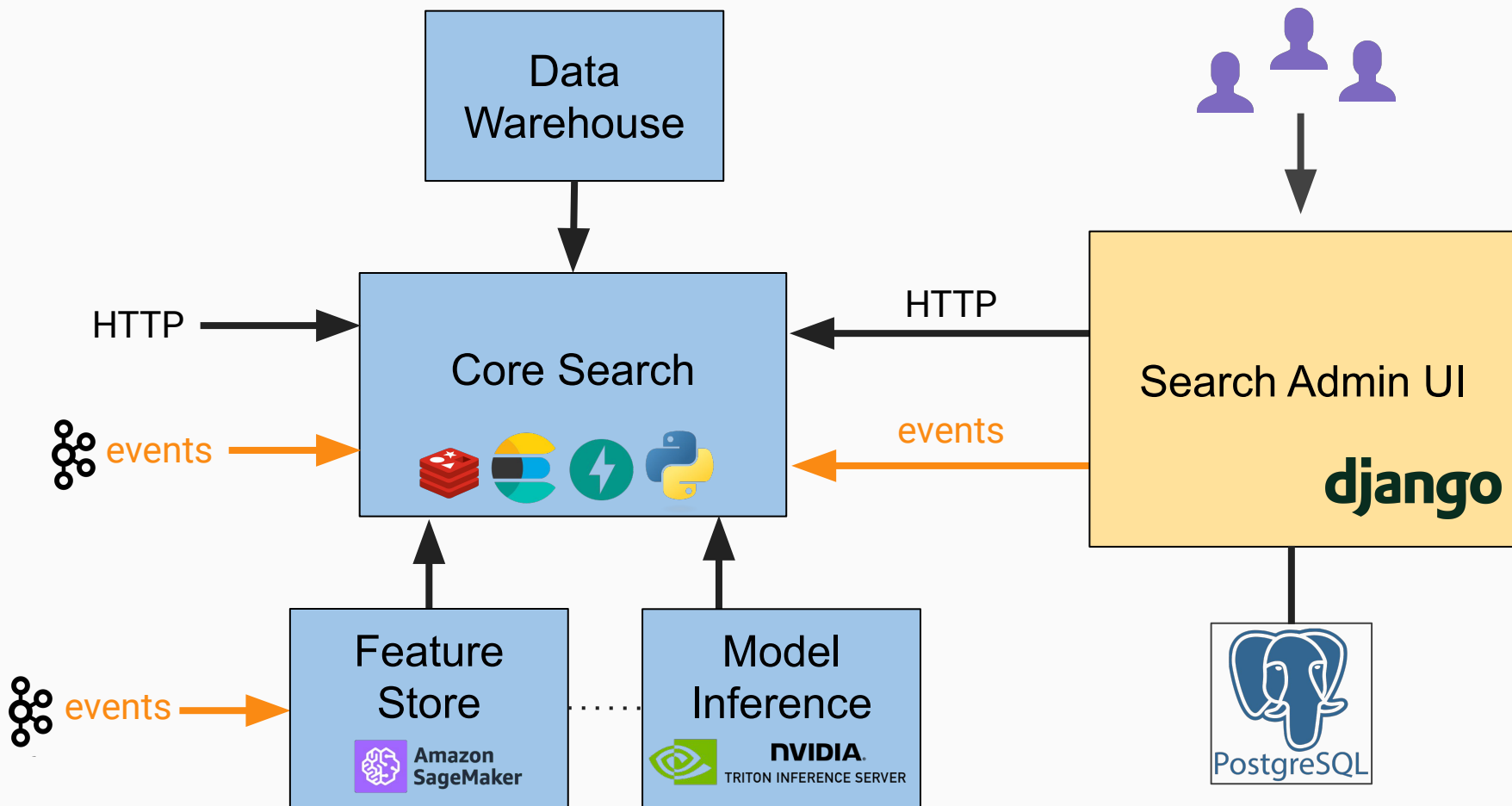
**Side note: By 2022, many platform teams** (but that's a story for another time...)

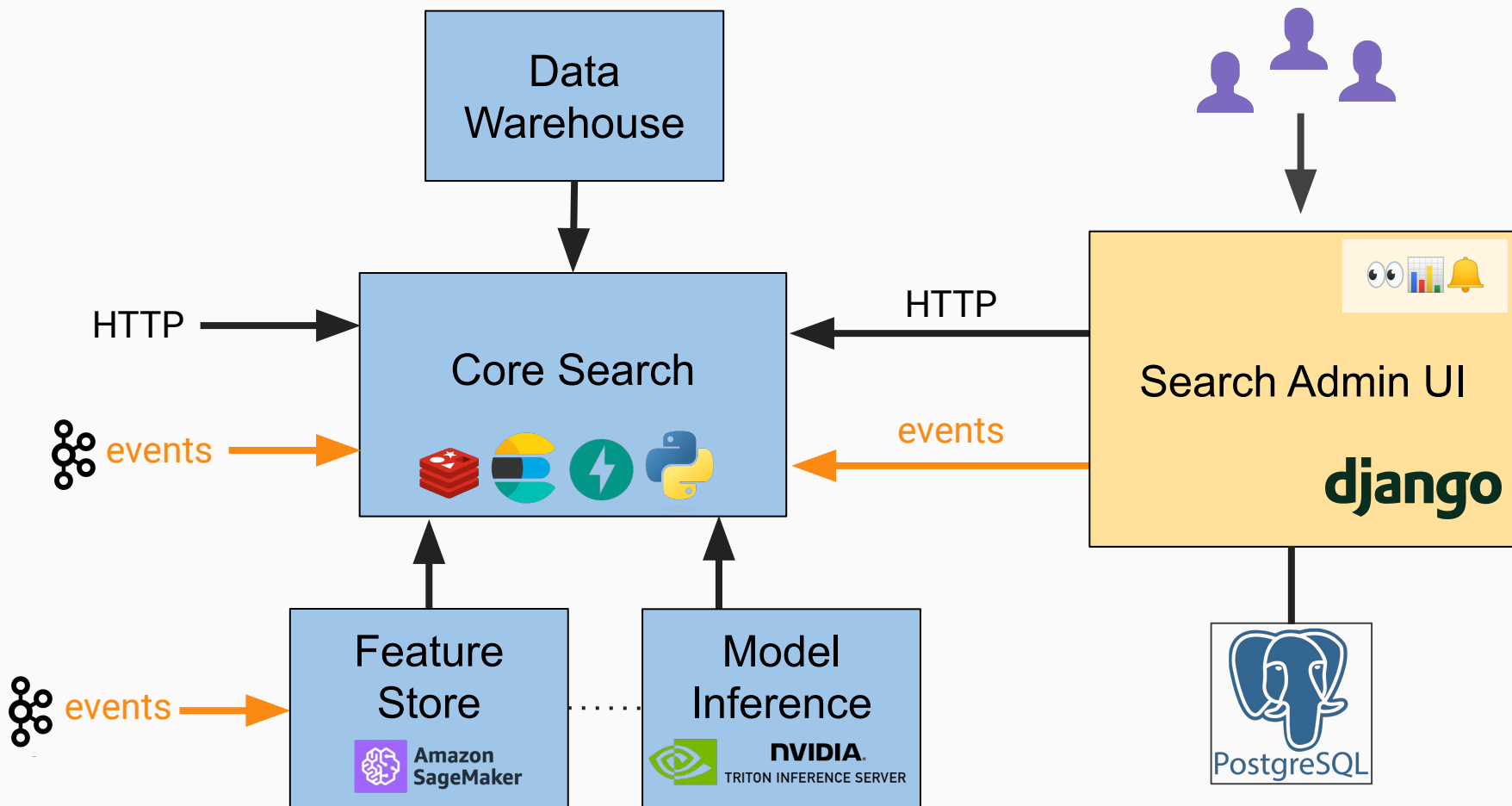












# The end goal

"Search teams and search engineers will **identify impactful opportunities** to optimise search, **build solutions** to solve those problems, and **rigorously validate** the impact."

And a few key ingredients 🍍 🍏 🍇 that helped us get there...

**Responsibility and ownership** over the search experience  
(including **relevance**)



**Autonomy and empowerment** to build solutions  
(while managing cognitive load)



TAGGED IN

## Search



Source Diving

Developer blog from the folks at Cookpad.  
@cookpad\_dev on twitter

[More information](#)

FOLLOWERS

610

ELSEWHERE



MORE, ON MEDIUM



Muhammad Hammad Khan in Source Diving

Dec 12, 2022 · 18 min read



## Learning to Boost—Query-time relevance signal boosting @ Cookpad

[Read more...](#)

174



Takayuki Watanabe in Source Diving

Jul 23, 2021 · 9 min read



Terraform repository



Atlantis

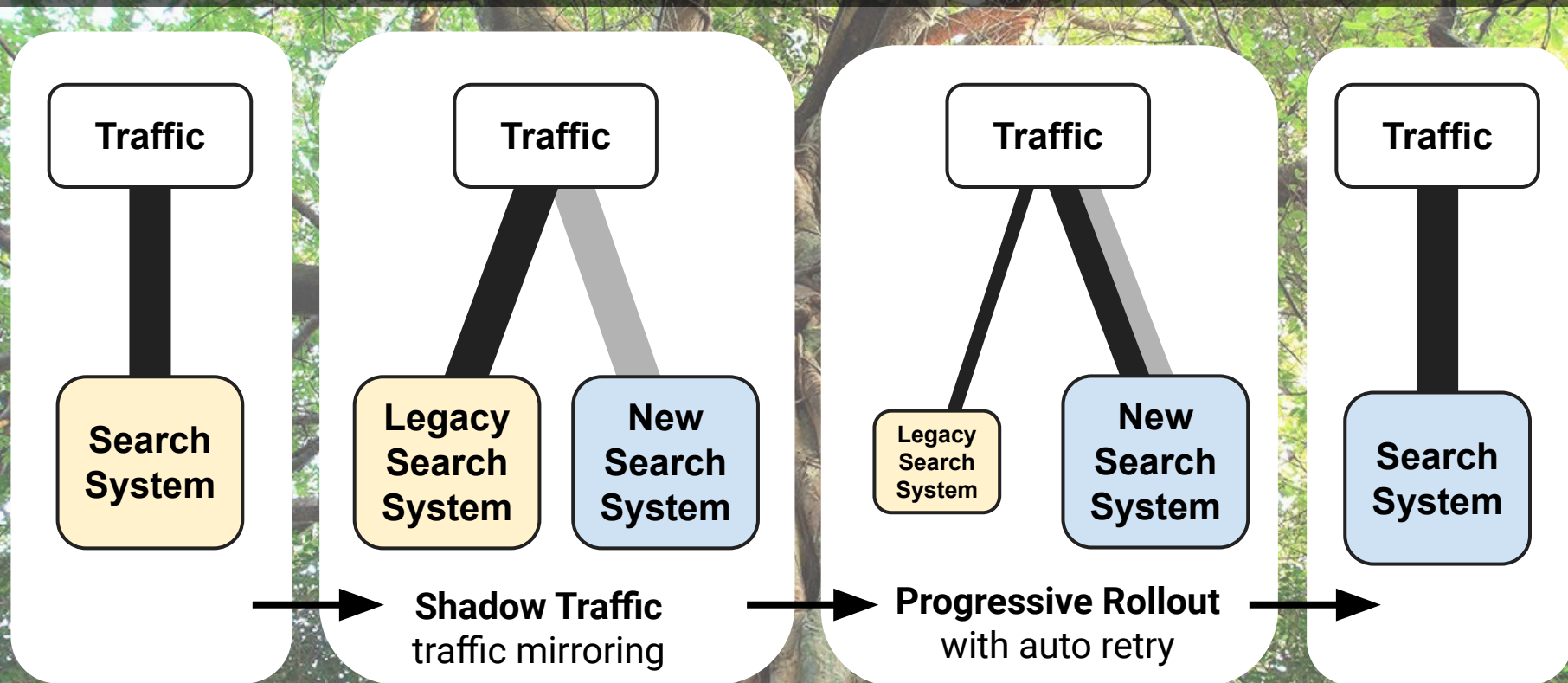


Source Diving → <https://sourcediving.com>

# Strangler fig for system migration (and putting the practices together)



# Strangler fig for system migration (and putting the practices together)



Search at Cookpad: building new infrastructure  
<https://sourcediving.com/search-at-cookpad-building-new-infrastructure-dc58f4eab93f>

 <https://www.science.org/content/article/being-strangled-may-save-tree-s-life>



# Cooking up a new search system

Recipe search at Cookpad

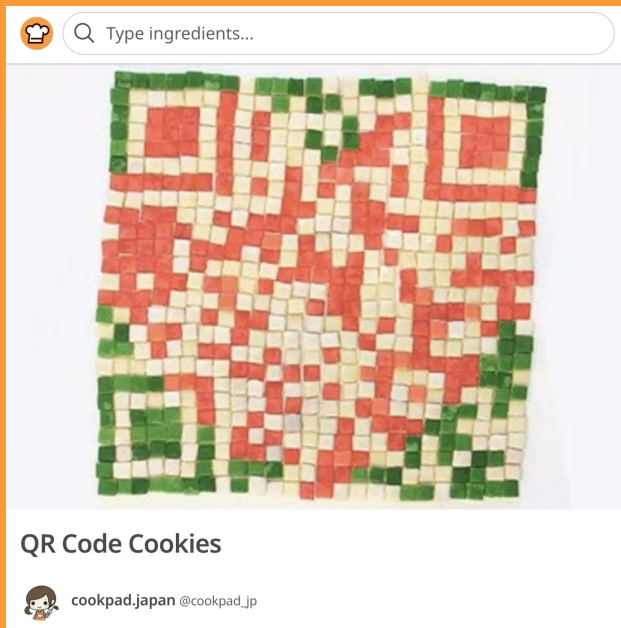
**Berlin Buzzwords**

19 June 2023

**Matt Williams**

Principal Engineer – Search and Discovery  
at **Cookpad Global**

**Web:** [mattjw.net](http://mattjw.net) | **Social:** @voxmjw | **Cookpad:** @matt



<https://cookpad.com/uk/recipes/146480-qr-code-cookies>

# Thank you!

Web: [cookpad.com](https://cookpad.com)

Android

iOS

Engineering blog  
<https://sourcediving.com>

Go Search v2  
<https://speakerdeck.com/giga811/go-global-search-2>

