

# Deploying differential privacy at Wikimedia with Tumult Labs

**Webinar 4 for Wikimedia Foundation, July 2022**

**Michael Hay  
Tumult Lab**

Tumult and Tumult Labs are trademarks of Tumult Labs, Inc.

# Recap, outline of Webinar 4

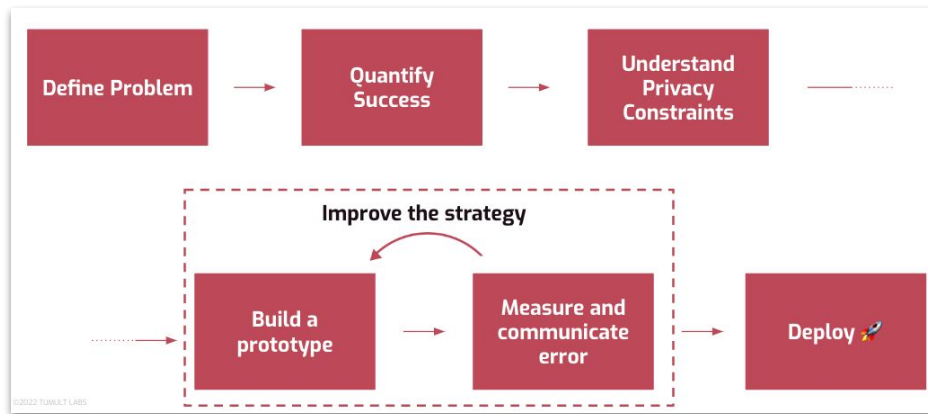
- Differential privacy (DP) requires changes to data analysis
  - To find desirable privacy-utility trade-off, iteration and optimization are key
- Previously: using Tumult Analytics to optimize the trade-off, and get useful results
- Today: an overview of critical steps in the deployment process

# Recap, outline of Webinar 4

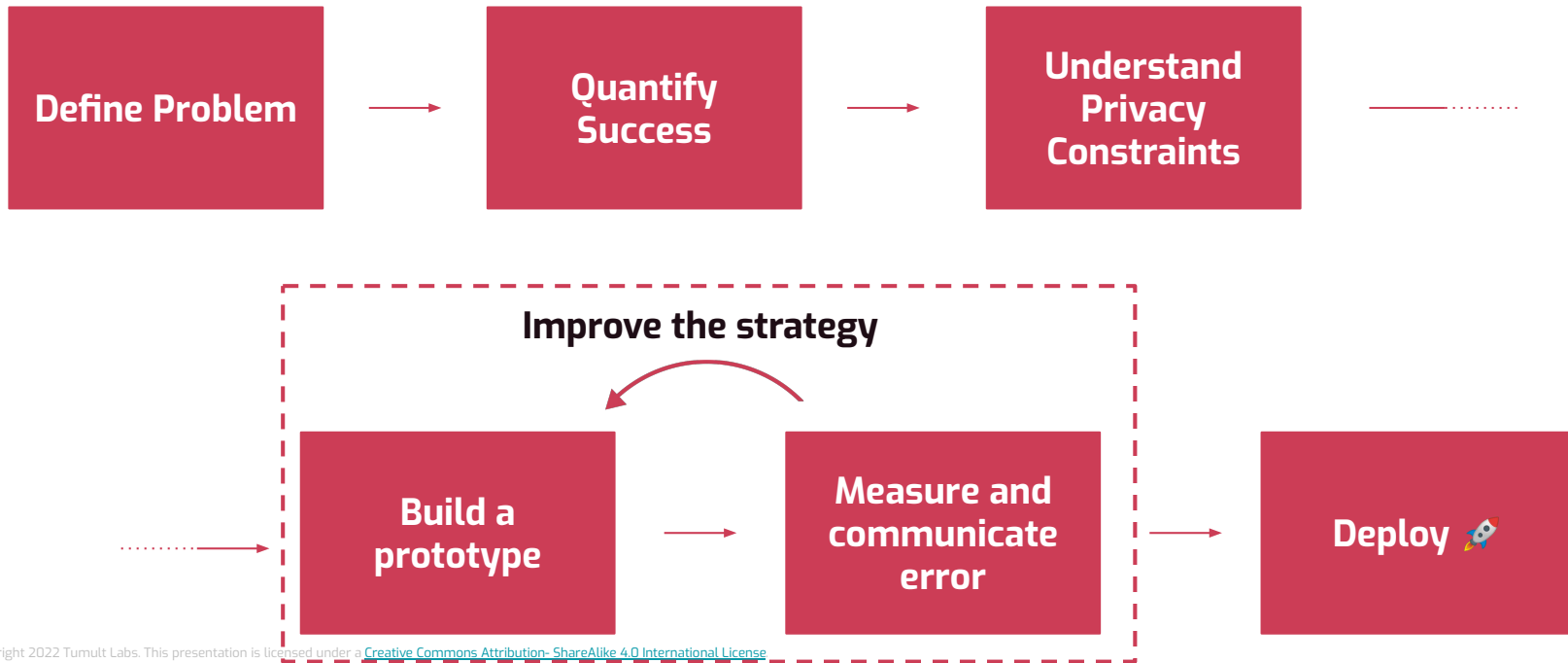
1. Walk through 6 stages in the deployment process

2. Use country-project-page histogram as a case study

3. Group activity: brainstorm another potential DP data release at WMF



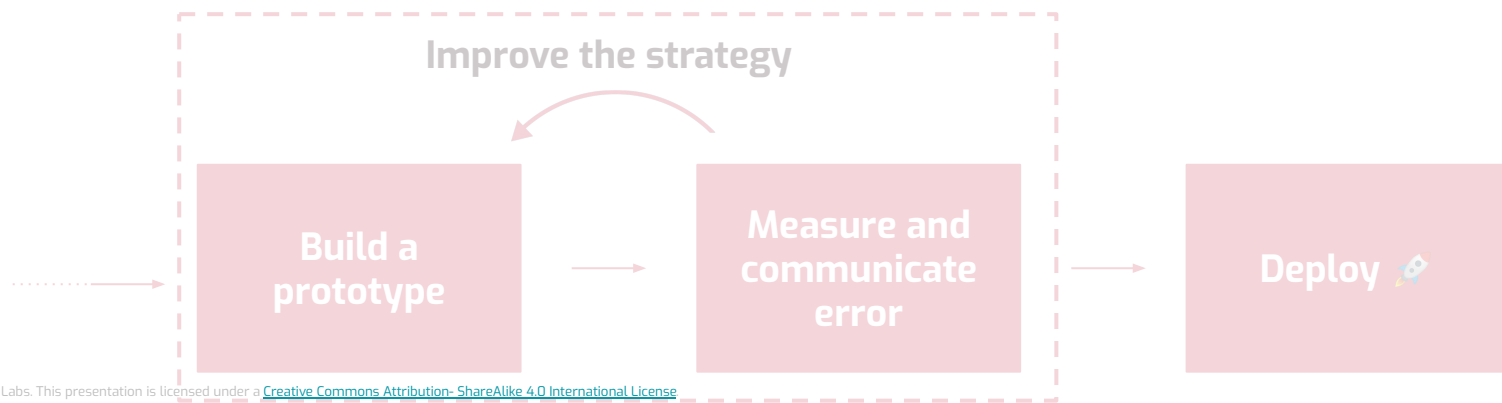
# DP Deployment Process



# DP Deployment Process

## Define Problem

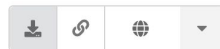
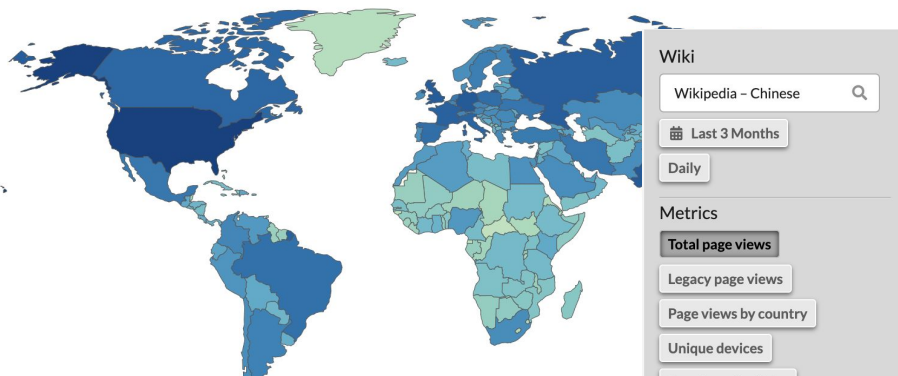
- *What are you trying to achieve?*
- *List key stakeholders, talk with them, understand use cases*
- **Outcome:** *technical specification, incl. input/output schemas*



# Pilot: views per article per country

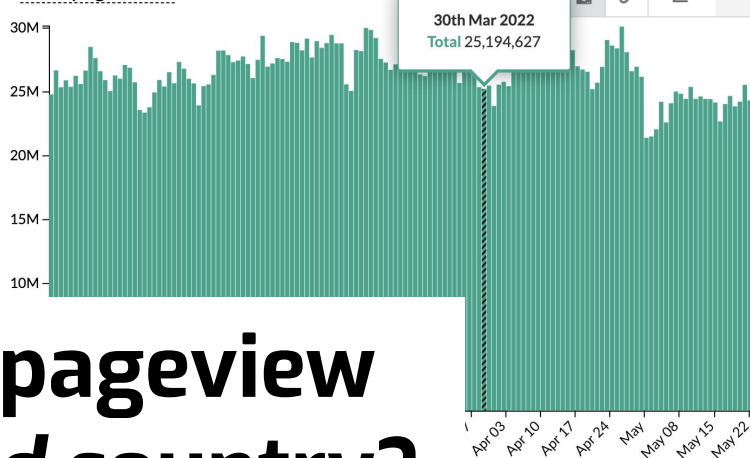
Page views by country

<https://stats.wikimedia.org/#/all-projects/reading/page-views-by-country/>



<https://stats.wikimedia.org/#/all-projects/reading/total-page-views/>

Total page views



## Can we use DP to release pageview counts by both project *and* country?

Total: 40

# Pilot: views per article per country

Why is this a *useful* problem to solve?

- Disaggregate trends within languages that are spoken in many countries
  - Spanish, English, Arabic, Vietnamese, Chinese, etc.
- Largest (and most unwieldy) dataset that WMF has
  - If we can successfully do it here, we can do it anywhere

Why is this a *difficult* problem to solve?

- Many country-project combos identify small user groups
- High cost of failure: censorship, sensitive topics, unmasking of editors, etc.
- Tension between *data minimization* and *differential privacy* (more later)

# DATA

date	project	page_id	page_title	actor_signature	country
July 27, 2022	English	123	"Differential Privacy"	0x456FD4A56E	USA
...	...	...	...	...	...

easily derived from  
`wmf.pageview_actor`

≈ 500M rows / day

# QUERY ≈

```
data.filter("date == {today}")  
  .groupby("page_id", "project", "country")  
  .count()
```

# RESULT

project	page_id	country	count
English	123	USA	9,451
English	123	Cameroon	1
...	...	...	...

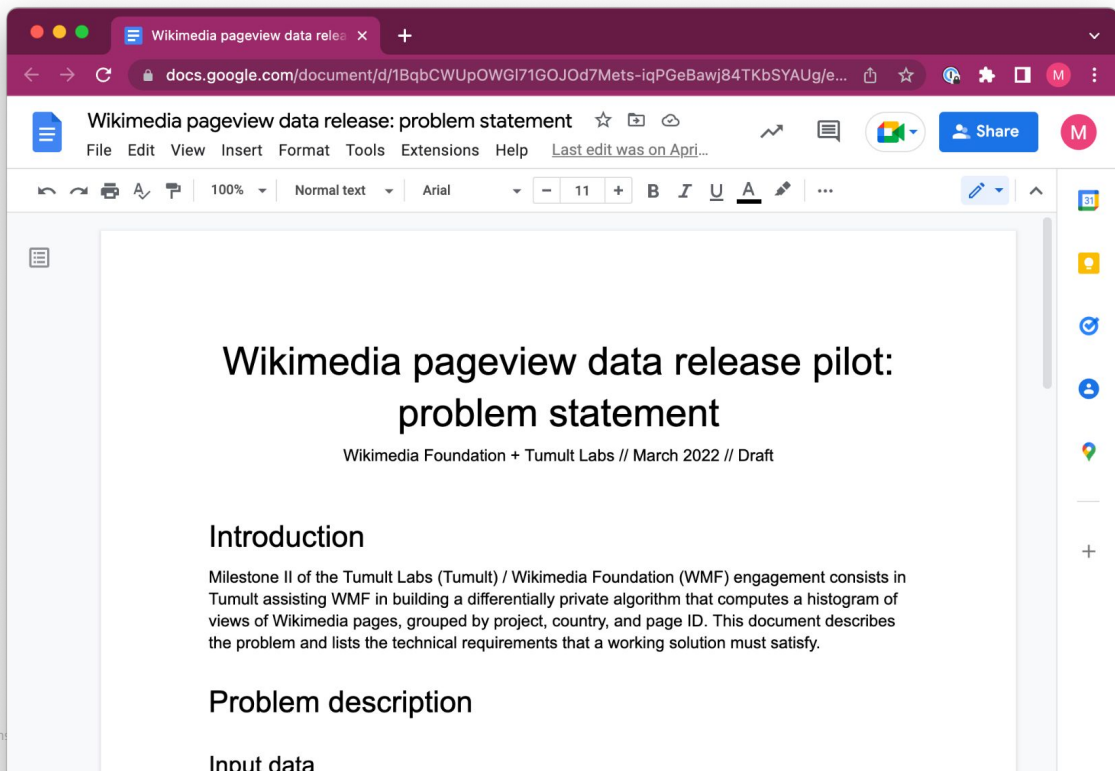


# Pilot: views per article per country

- Result is sparse
  - Majority of pages receive 0 or 1 pageview / day
- Counts will be noisy (because differential privacy)
- Small counts therefore unreliable

⇒ Only release counts that are above specified threshold

# Pilot: views per article per country



The image shows a screenshot of a Google Docs document. The browser's address bar shows the document ID: docs.google.com/document/d/1BqbCWUpOWGI71GOJ0d7Mets-iqPGeBawj84TKbSYAUg/e... The document title is "Wikimedia pageview data release: problem statement". The document content includes a main title, a subtitle, and an introduction paragraph.

## Wikimedia pageview data release pilot: problem statement

Wikimedia Foundation + Tumult Labs // March 2022 // Draft

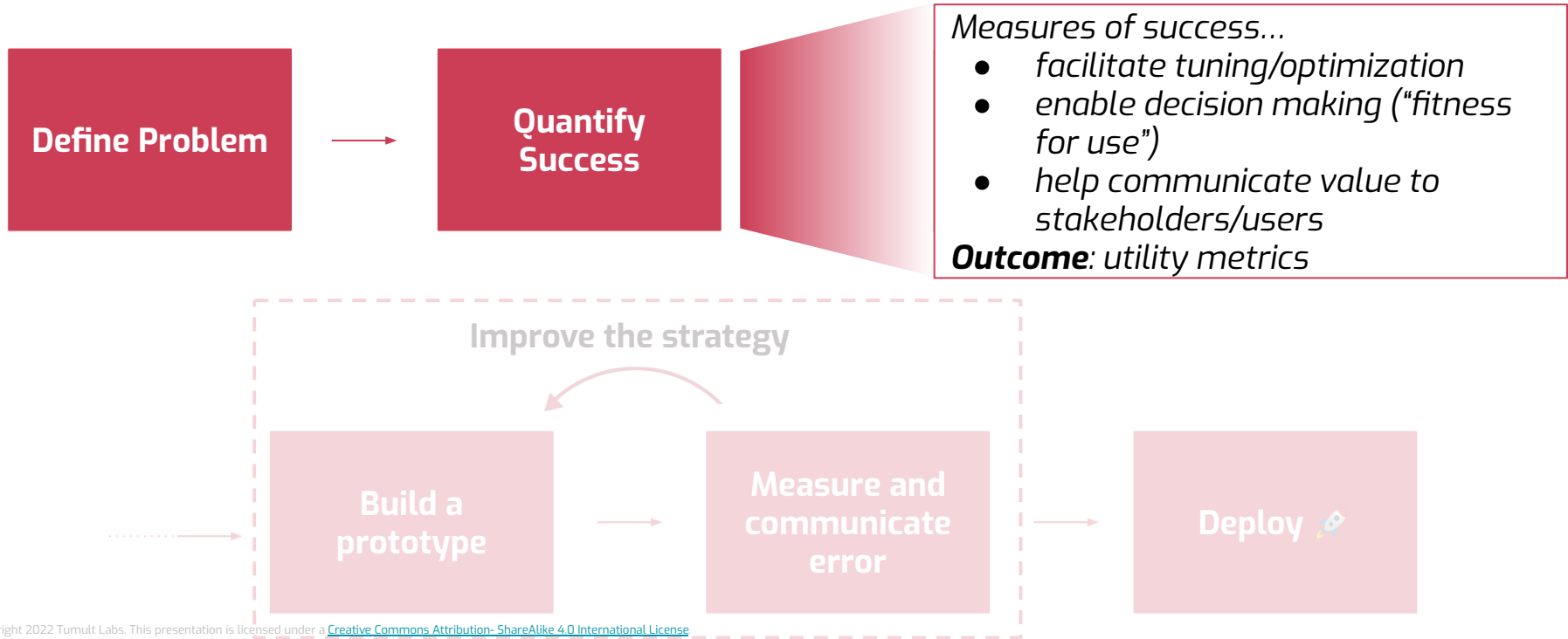
### Introduction

Milestone II of the Tumult Labs (Tumult) / Wikimedia Foundation (WMF) engagement consists in Tumult assisting WMF in building a differentially private algorithm that computes a histogram of views of Wikimedia pages, grouped by project, country, and page ID. This document describes the problem and lists the technical requirements that a working solution must satisfy.

### Problem description

### Input data

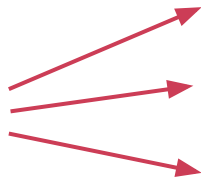
# DP Deployment Process



# Pilot: utility metrics

Two primary utility goals:

1. Avoid publishing *misleading* data



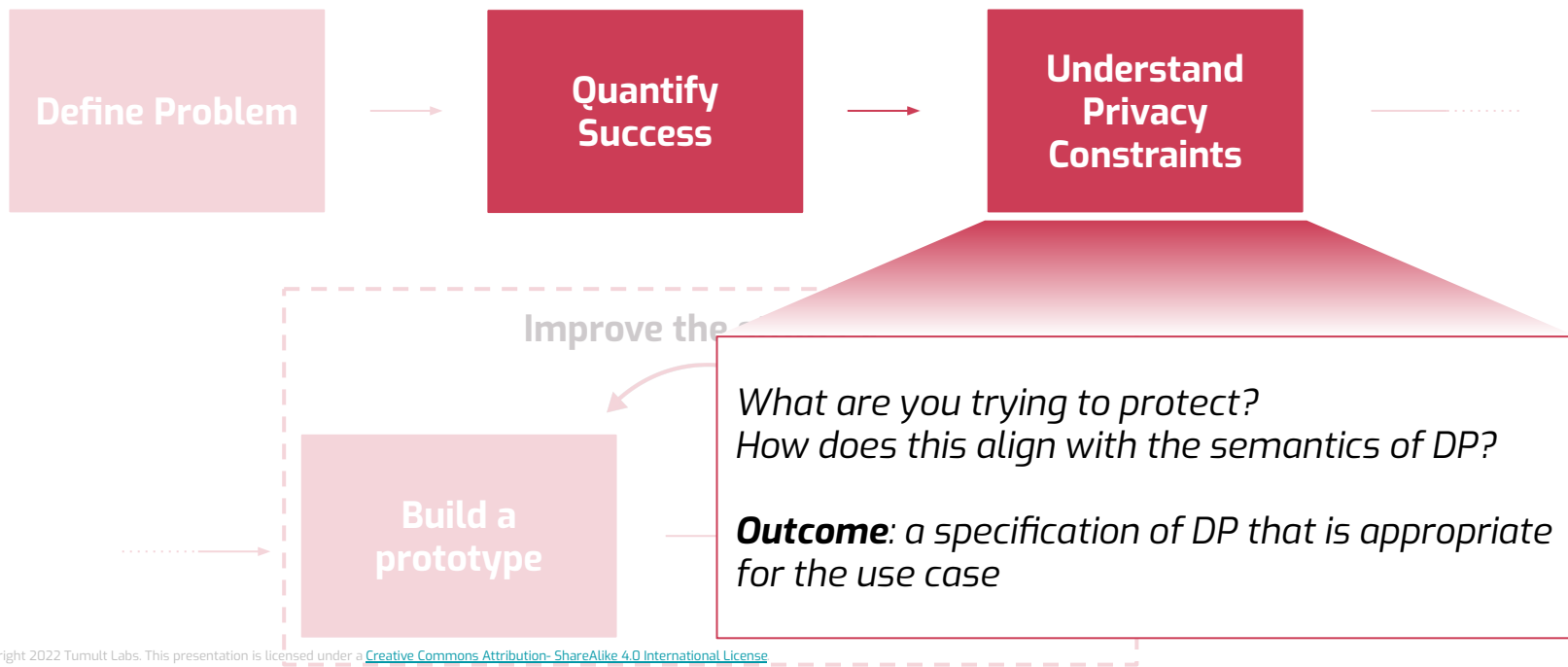
**Metrics**  
Median relative error  
% counts with relative error  $\leq 0.1$   
% counts with relative error  $\leq 0.25$   
% counts with relative error  $\leq 0.5$   
Spurious rate

2. Release as much data as possible  
(subject to privacy constraints)







Number of counts released  
Drop rate

# DP Deployment Process



# Pilot: understanding privacy constraints

What is being protected?	Meaningful Privacy	Ease of Implementation
A single row in the input		
All rows associated with any single user (on the given day)		

# Pilot: problems with actor signature

What is a “user”?

**ActorSignature = MD5(IP, UserAgent)**



*Failure 1: One user, many signatures*

IP address changes while browsing ⇒  
signature changes as well.

Problem for areas where most  
browsing happens on mobile (India,  
Indonesia, Mexico, etc.)

*Failure 2: Many users, one signature*

Many users have same IP and UA ⇒ all  
hash to the same signature

Problem for browsing within institutions  
where people might all have the same  
devices (universities, offices, etc.)

# Pilot: client-side filtering

Cookie attached to web request that indicates whether to include in DP aggregations (`include=Y`).

Client-side “filtering”: only the first  $T$  requests (for distinct pages) will be marked for inclusion (the rest have `include=N`).

At server, when doing DP aggregations,

- only include those with `include=Y`
- initialize `tmlt.analytics.Session` to protect up to  $T$  rows (for distinct pages)



# Pilot: client-side filtering

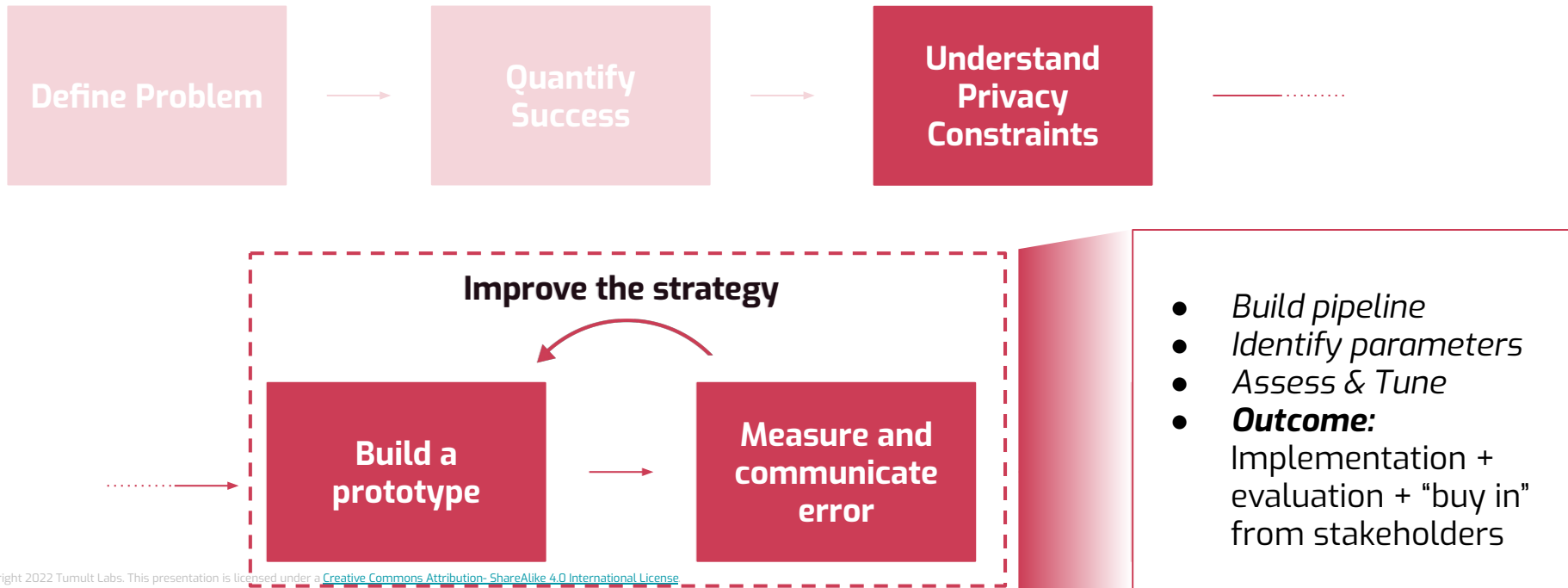
*Failure 1: One user, many signatures*

Stability > ActorSignature, because cookies are cleared and browser changes less than IP address changes

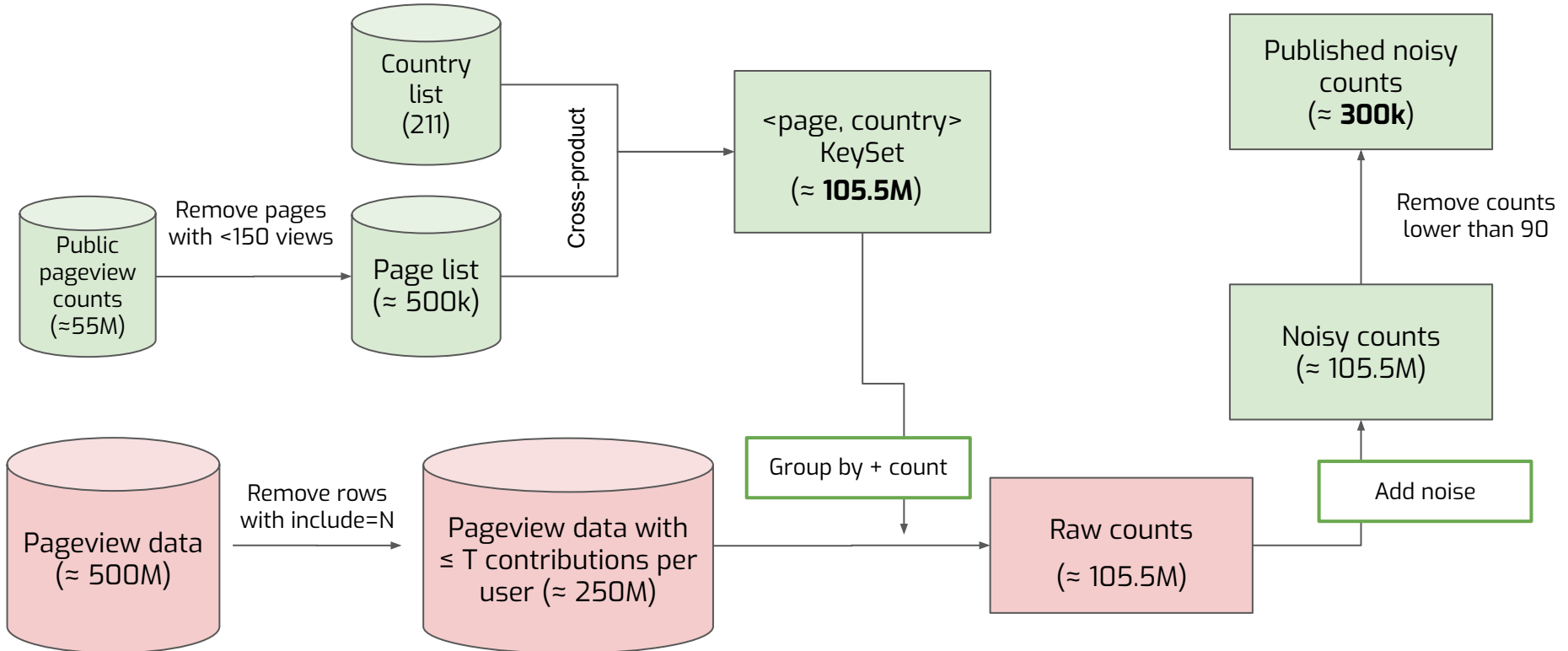
*Failure 2: Many users, one signature*

Disaggregation is possible, because distinct devices will all say to include their first  $T$  pages.

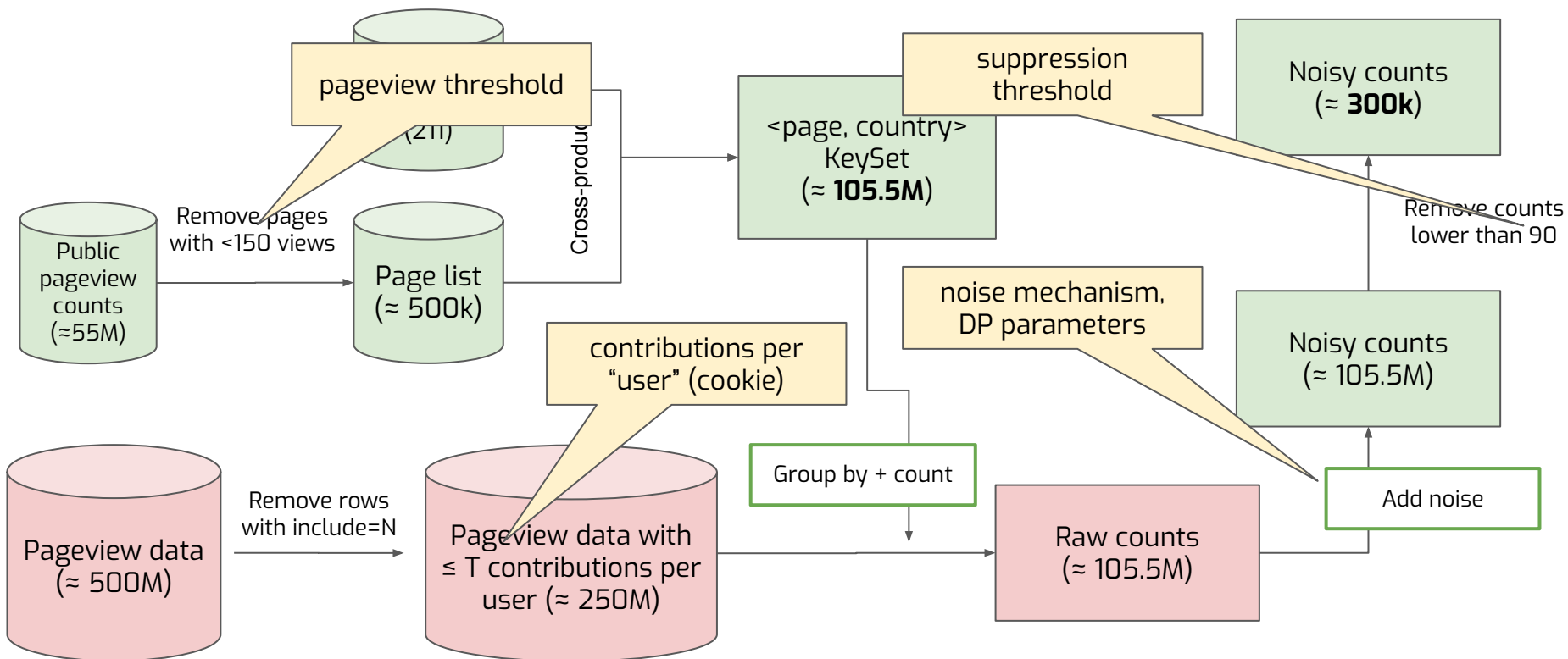
# DP Deployment Process



# Pilot: pipeline



# Pilot: identify parameters

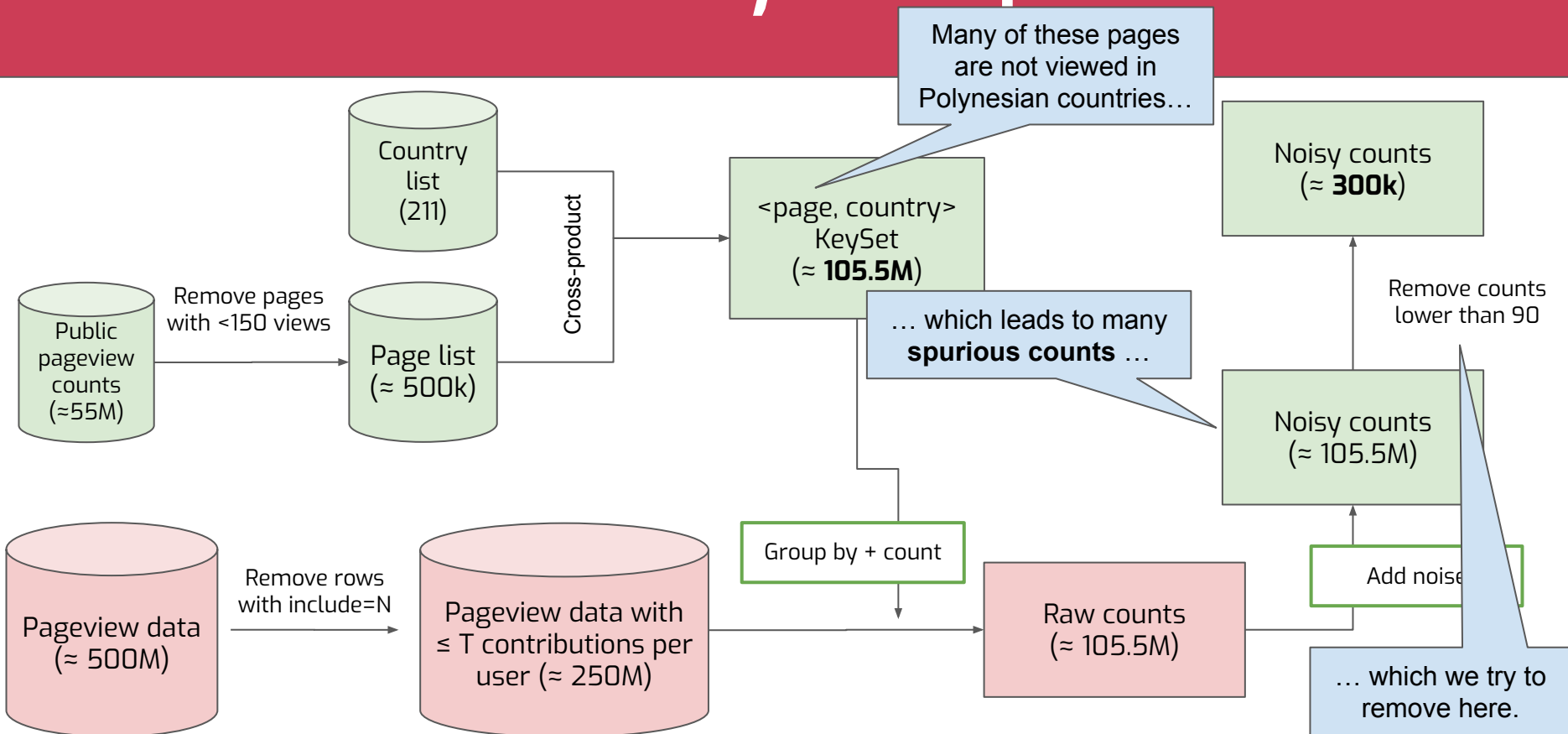


# Pilot: evaluation & surprises

Geo region	Total published	Median relative error	Spurious Rate
...	...	...	...
Western_Europe	44,548	3.85%	0.18%
...	...	...	...
Polynesia	303	736.36%	99.39%
...	...	...	...



# Pilot: the “Polynesia” problem

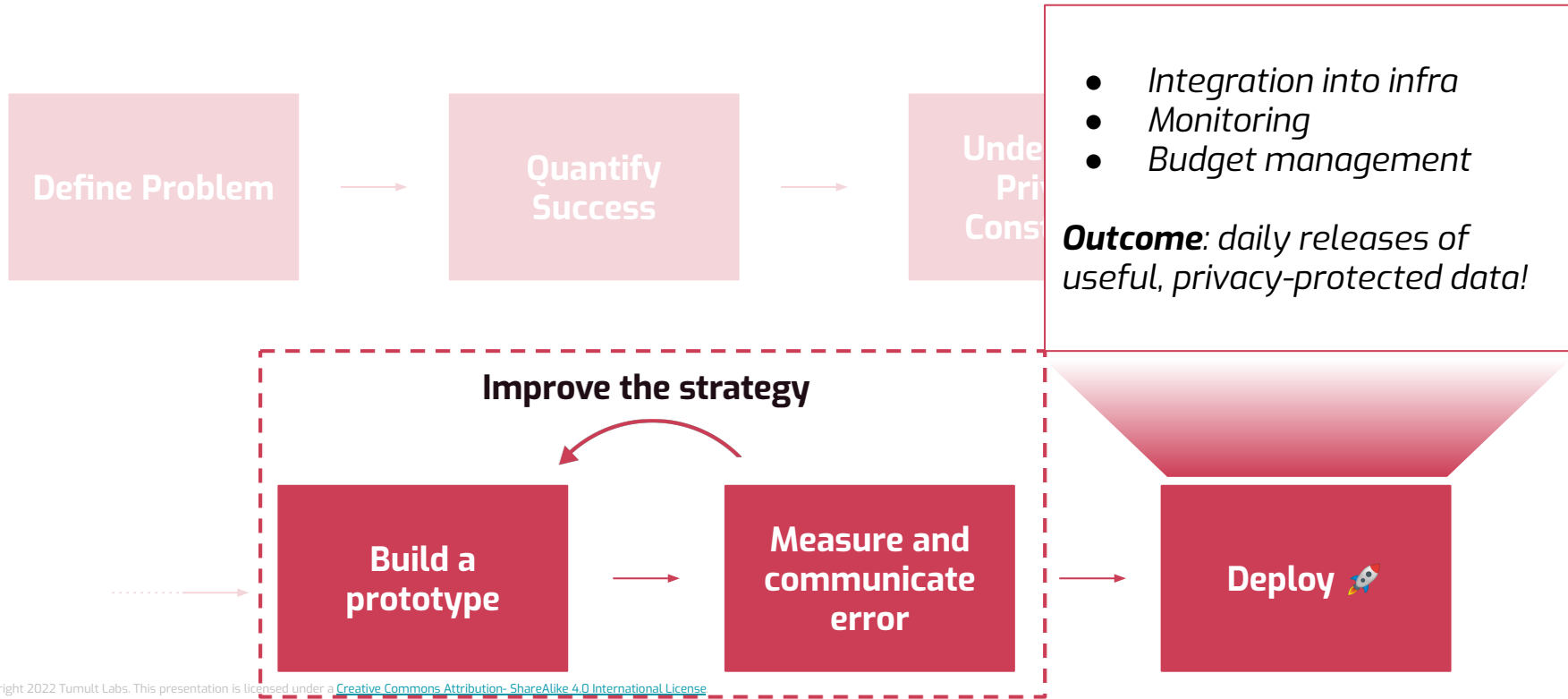


# Pilot: improved prototype

By using a different noise distribution, we greatly reduced the spurious rate (at the cost of a modest increase in relative error).

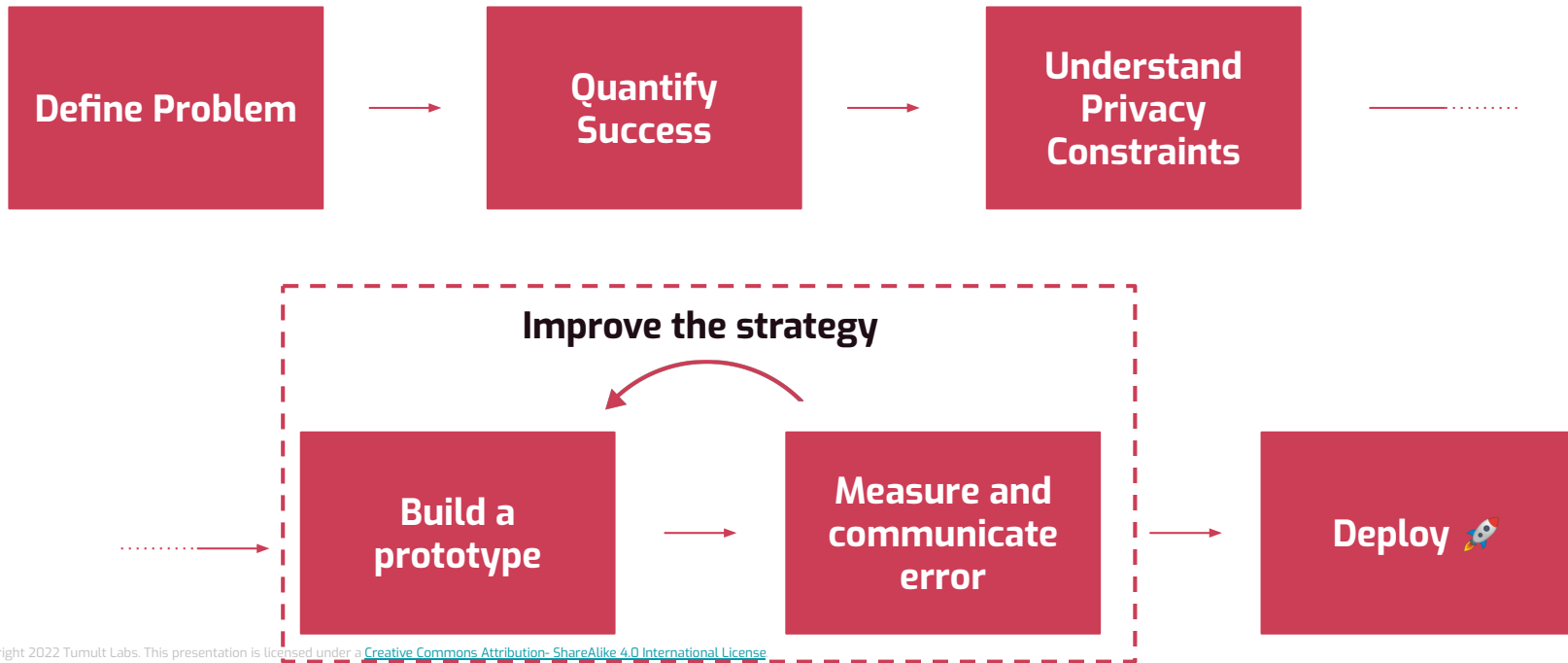
Geo region	Total published	Median relative error	Spurious Rate
...	...	...	...
Western_Europe	44,277	6.98%	0.00%
...	...	...	...
Polynesia	0	N/A	N/A
...	...	...	...

# DP Deployment Process



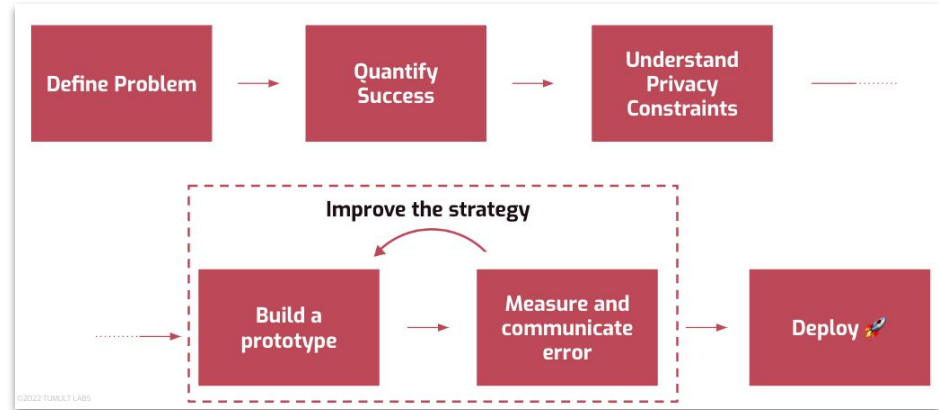


# DP Deployment Process



# Conclusion

- Reviewed 6 stages in the deployment process
- Used country-project-page histogram as a case study
- **Up next:** Group activity!
  - Brainstorm another potential DP data release at WMF



# Group Activity!

Let's talk about possible use cases, and pick one as a group.

Explore answers to these questions:

1. Who would use the data? To what purpose?
2. How sensitive is the data? What would we want to protect in it?
3. How do we quantify success?
4. What do we expect to be challenging? What parameters would need tuning?

# Thank you! Questions?

Michael Hay  
michael@tmlt.io  
@michaelghay

Tumult and Tumult Labs are trademarks of Tumult Labs, Inc.