Speech Commands

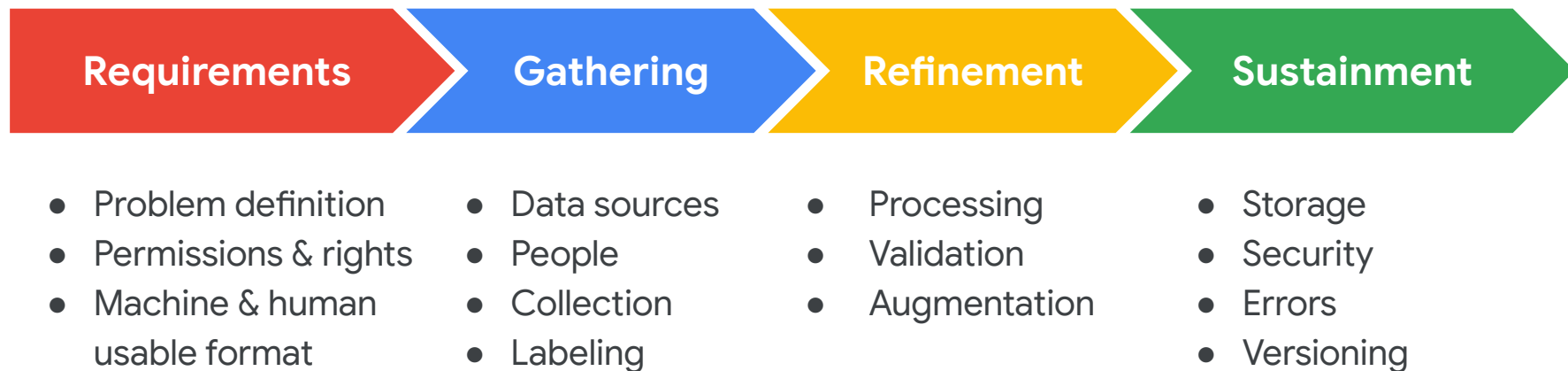# Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition
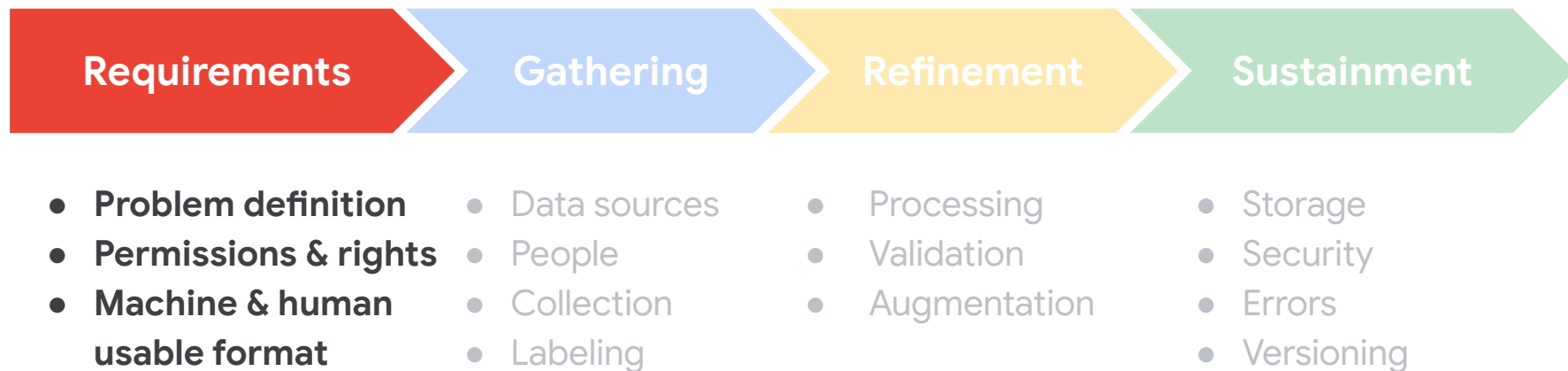
Pete Warden
Google Brain
Mountain View, California
petewarden@google.com

April 2018

Data Engineering

Requirements	Gathering	Refinement	Sustainment

Requirements
- Problem definition
- Permissions & rights
- Machine & human usable format

Gathering
- Data sources
- People
- Collection
- Labeling

Refinement
- Processing
- Validation
- Augmentation

Sustainment
- Storage
- Security
- Errors
- Versioning

Data Engineering

Requirements	Gathering	Refinement	Sustainment

- **Problem definition**
- **Permissions & rights**
- **Machine & human usable format**

- Data sources
- People
- Collection
- Labeling

- Processing
- Validation
- Augmentation

- Storage
- Security
- Errors
- Versioning

Requirements

- Need for a **public** dataset for keyword spotting (KWS)
- Previous datasets **not KWS-focused**
- Google, Apple, Amazon use **proprietary datasets**
- **Speech Commands**:
  - Permissively licensed
  - Research & commercial use ok
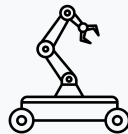- **Established new standard**

Requirements

- Need for a **public** dataset for keyword spotting (KWS)
- Previous datasets **not KWS-focused**
- Google, Apple, Amazon use **proprietary datasets**
- **Speech Commands**:
  - Permissively licensed
  - Research & commercial use ok
- **Established new standard**

Requirements

"yes" ✔
"no" ✘

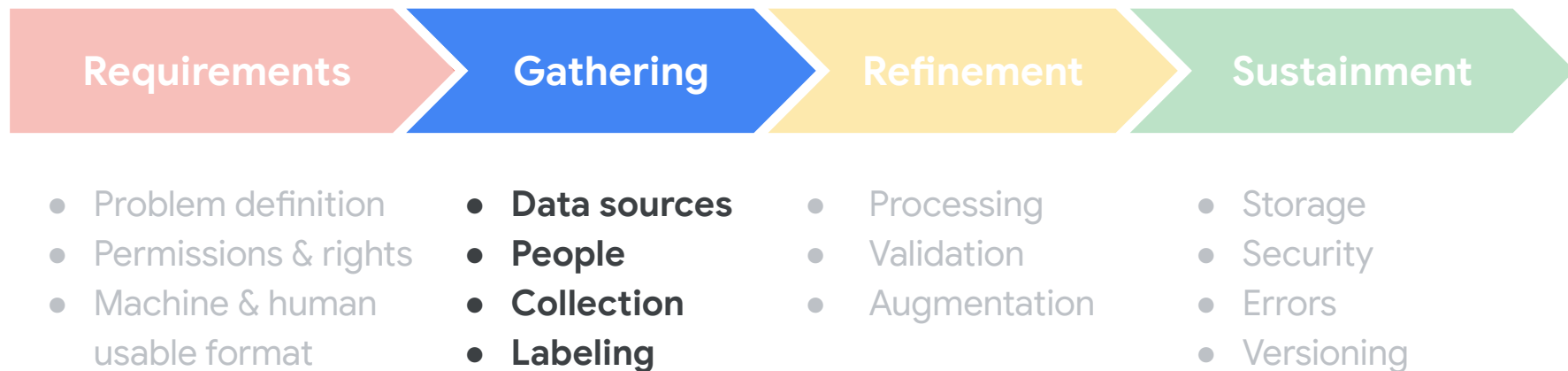**Common Use**

"left"
"right"
"go"
"stop"

Robotics

"one"
"two"
"four"
"six"

#

Numbers

Data Engineering

Requirements	Gathering	Refinement	Sustainment

- Problem definition
- Permissions & rights
- Machine & human usable format

- **Data sources**
- **People**
- **Collection**
- **Labeling**

- Processing
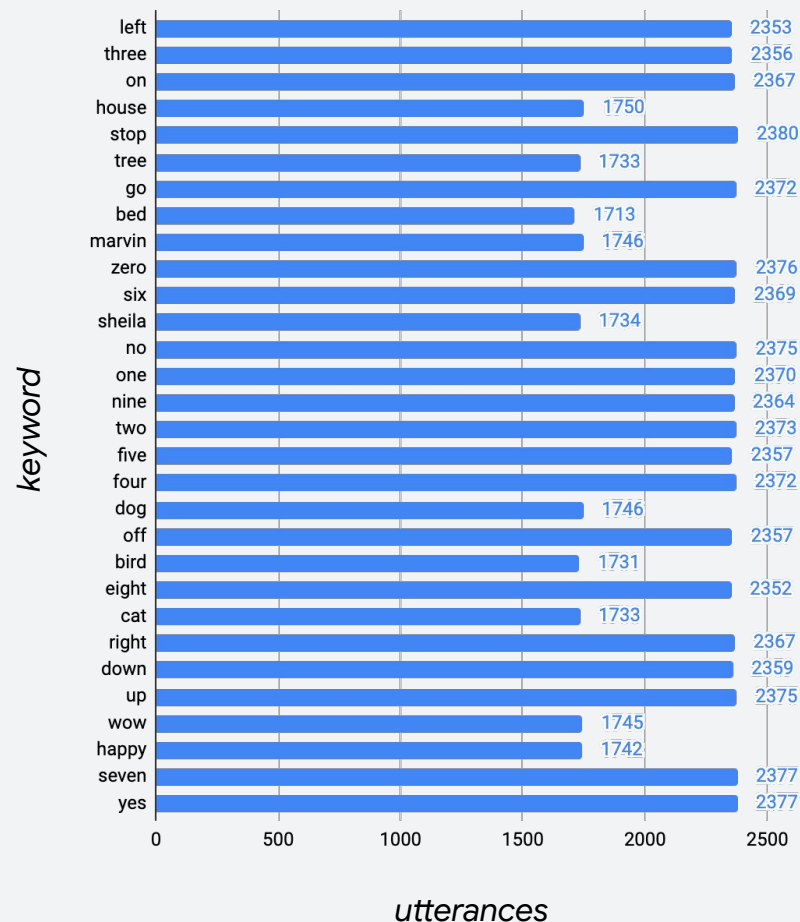- Validation
- Augmentation

- Storage
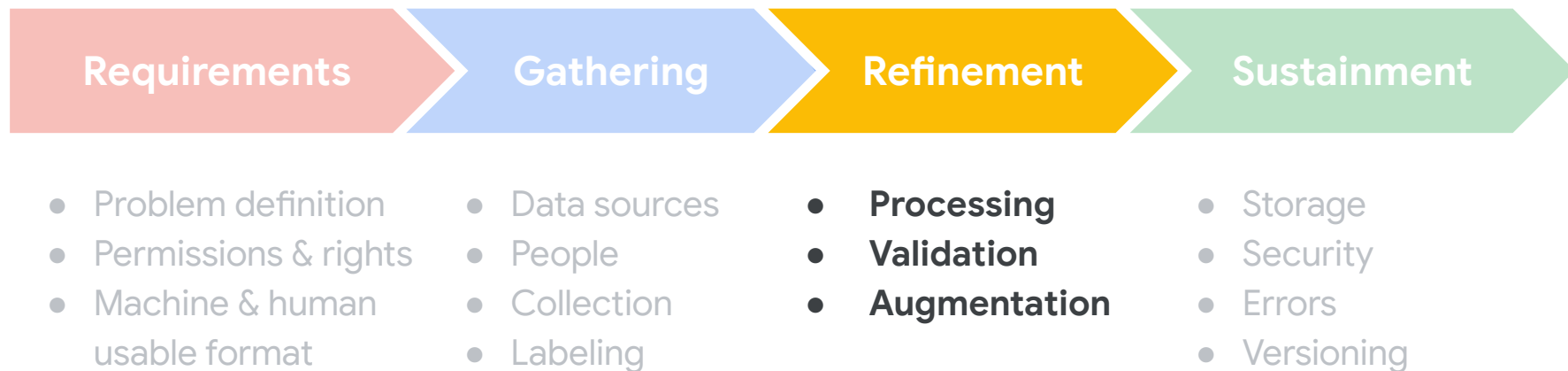- Security
- Errors
- Versioning

Data Collection

- **2,618** volunteers
  - consented to have their voices redistributed
  - Variety of accents
- > 1,000 examples for **each** keyword
- **Browser-based** (no app to install)

Data Engineering

Requirements	Gathering	Refinement	Sustainment

- Problem definition
- Permissions & rights
- Machine & human usable format

- Data sources
- People
- Collection
- Labeling

- **Processing**
- **Validation**
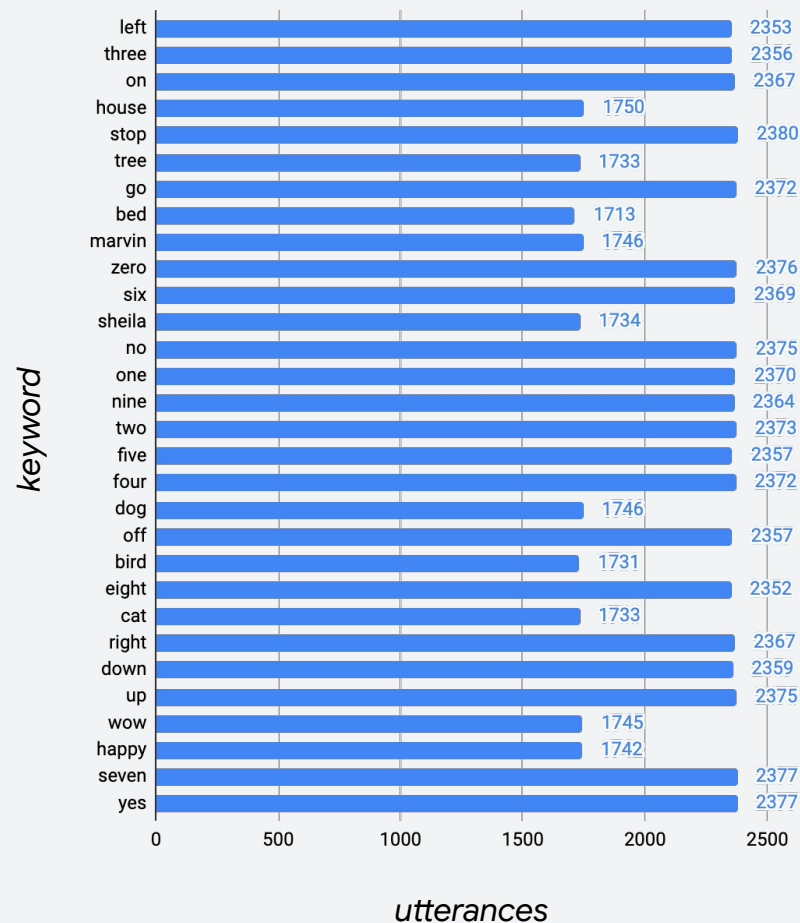- **Augmentation**

- Storage
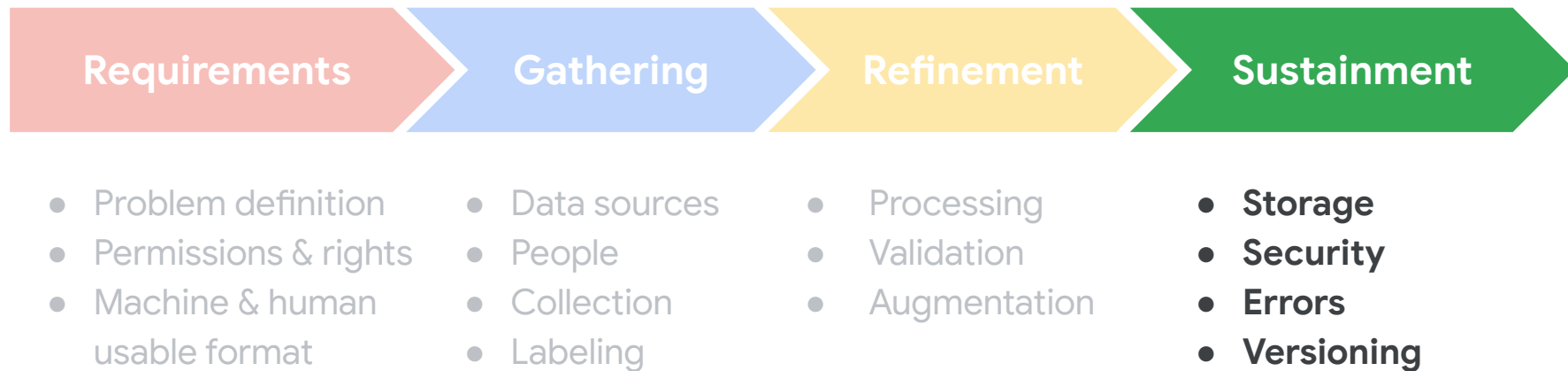- Security
- Errors
- Versioning

Data Validation

- Some data is **unusable**
  - Too quiet, wrong word, etc
- Started with **automated tools**
  - Remove low volume recordings
  - Extract loudest 1s (from 1.5sec examples)
- All 105,829 remaining utterances **manually reviewed** through crowdsourcing

keyword	utterances
left	2353
three	2356
on	2367
house	1750
stop	2380
tree	1733
go	2372
bed	1713
marvin	1746
zero	2376
six	2369
sheila	1734
no	2375
one	2370
nine	2364
two	2373
five	2357
four	2372
dog	1746
off	2357
bird	1731
eight	2352
cat	1733
right	2367
down	2359
up	2375
wow	1745
happy	1742
seven	2377
yes	2377

Data Engineering

Requirements	Gathering	Refinement	Sustainment

- Problem definition
- Permissions & rights
- Machine & human usable format

- Data sources
- People
- Collection
- Labeling

- Processing
- Validation
- Augmentation

- **Storage**
- **Security**
- **Errors**
- **Versioning**
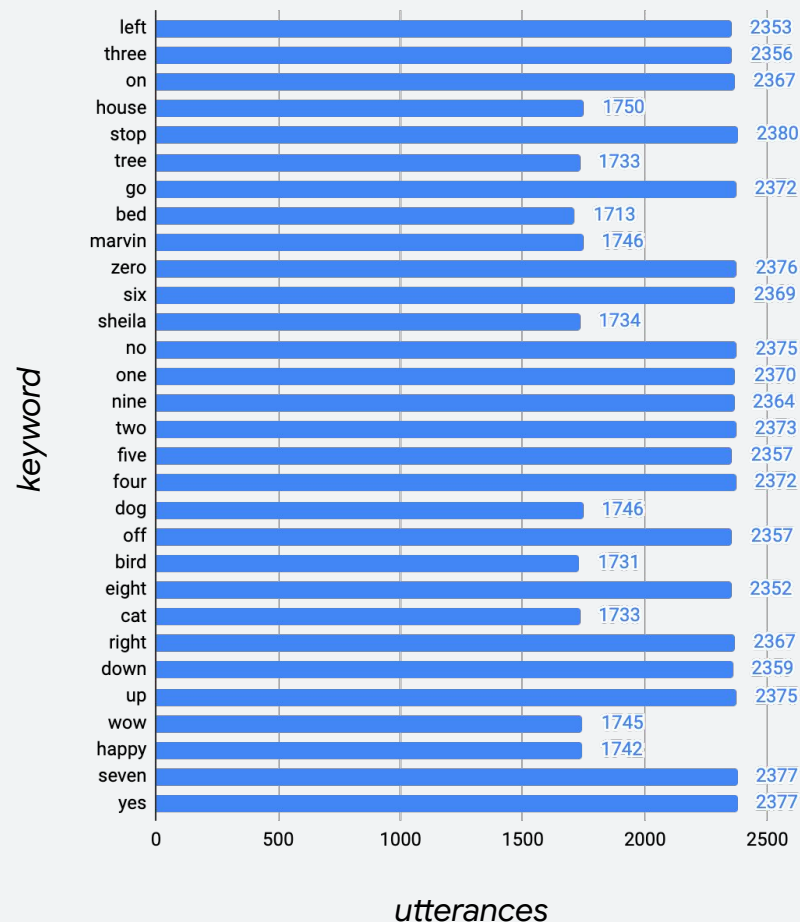
Sustaining KWS Research

- Speech Commands is now in **v2**
  - **Expanded to 35** keywords from original 10
- Includes train/validation/test splits
- Expand to **new languages**?

Why is **wide availability** of data *important*?

Compare	Without a standard dataset, no easy way to compare models

Benchmark	Good benchmarks: critical under *TinyML* constraints! Not just accuracy: power, memory, latency

Improve	Speech Commands: a standard training & evaluation dataset for each new KWS technique

Researchers using the same data: *"apples-to-apples" evaluation*