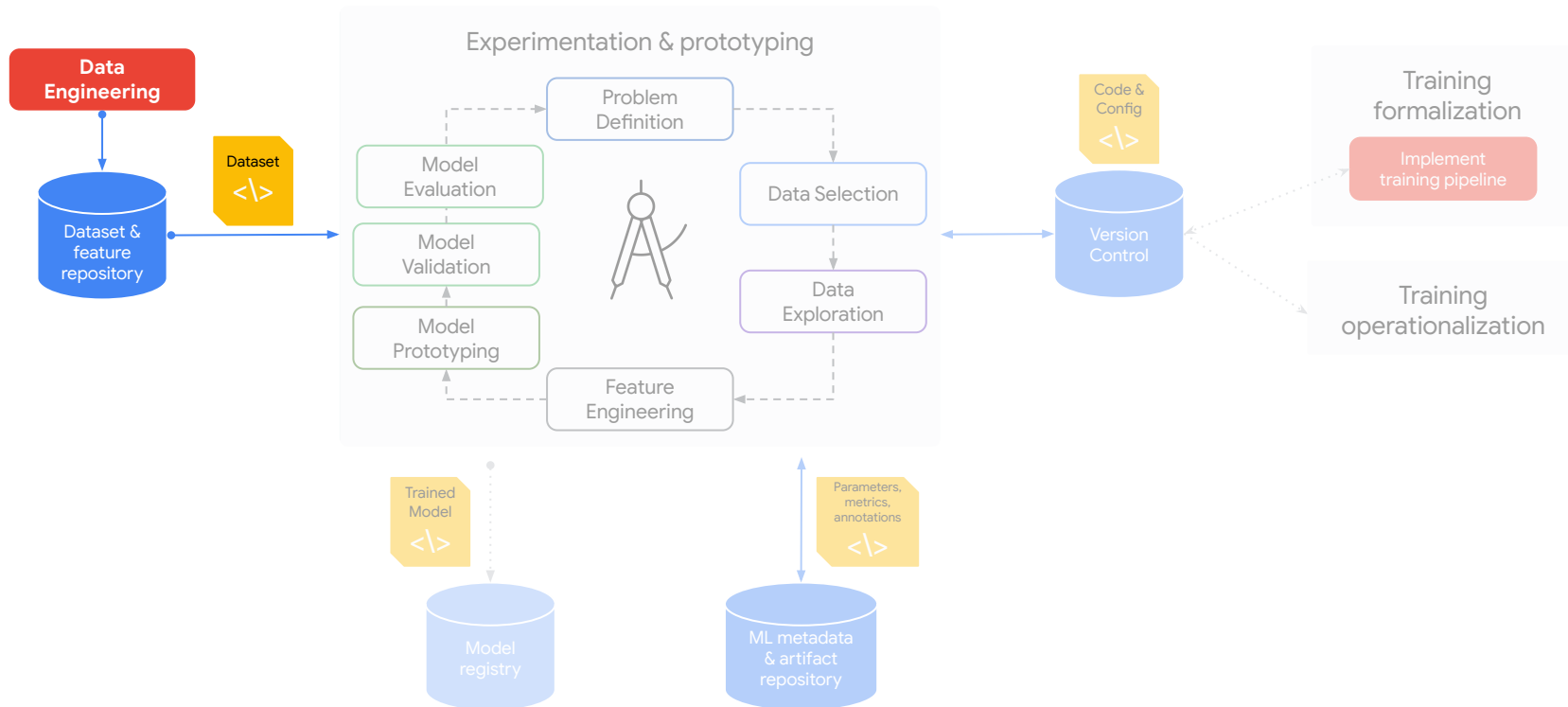


ML Development: Data Engineering



MLOps: ML Development



The MLOps Personas



ML
Engineer



ML
Researcher



Data
Scientist



Data
Engineer



Software
Engineer



DevOps



Business
Analyst

Data Engineering



Data Engineering



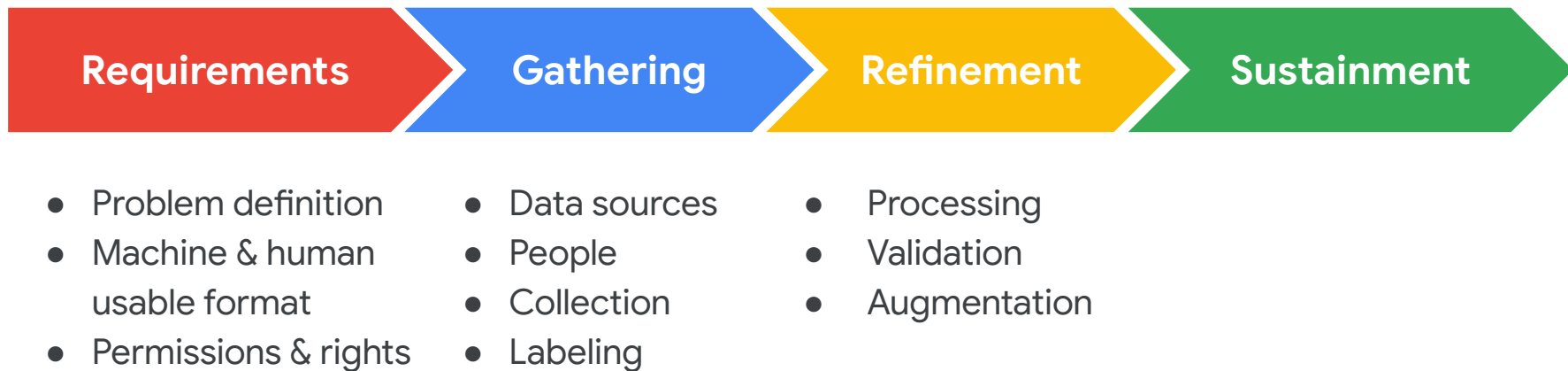
- Problem definition
- Machine & human usable format
- Permissions & rights

Data Engineering

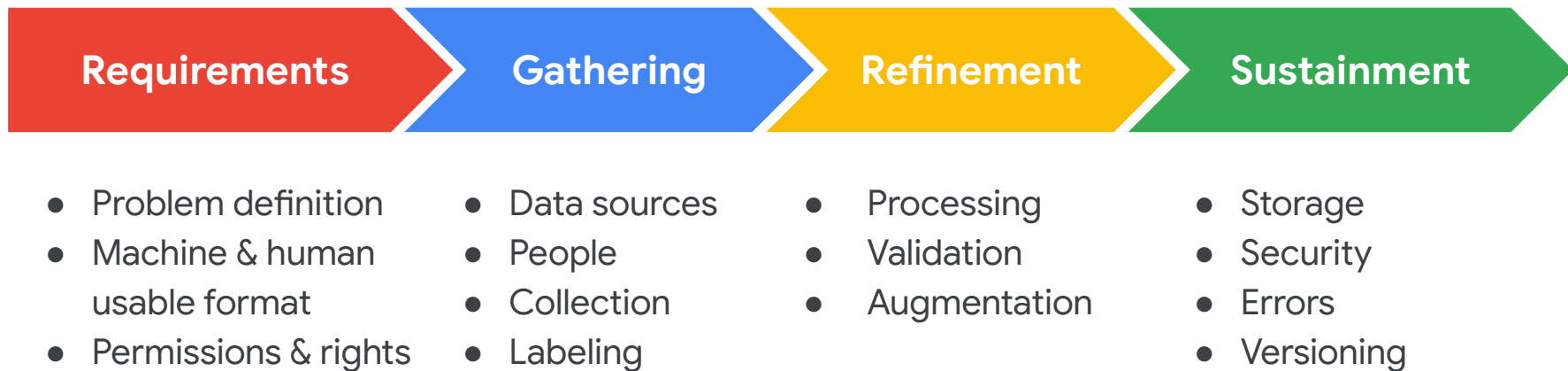


- Problem definition
- Machine & human usable format
- Permissions & rights
- Data sources
- People
- Collection
- Labeling

Data Engineering



Data Engineering



Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition

Pete Warden
Google Brain
Mountain View, California
`petewarden@google.com`

April 2018

Data Engineering



Data Engineering



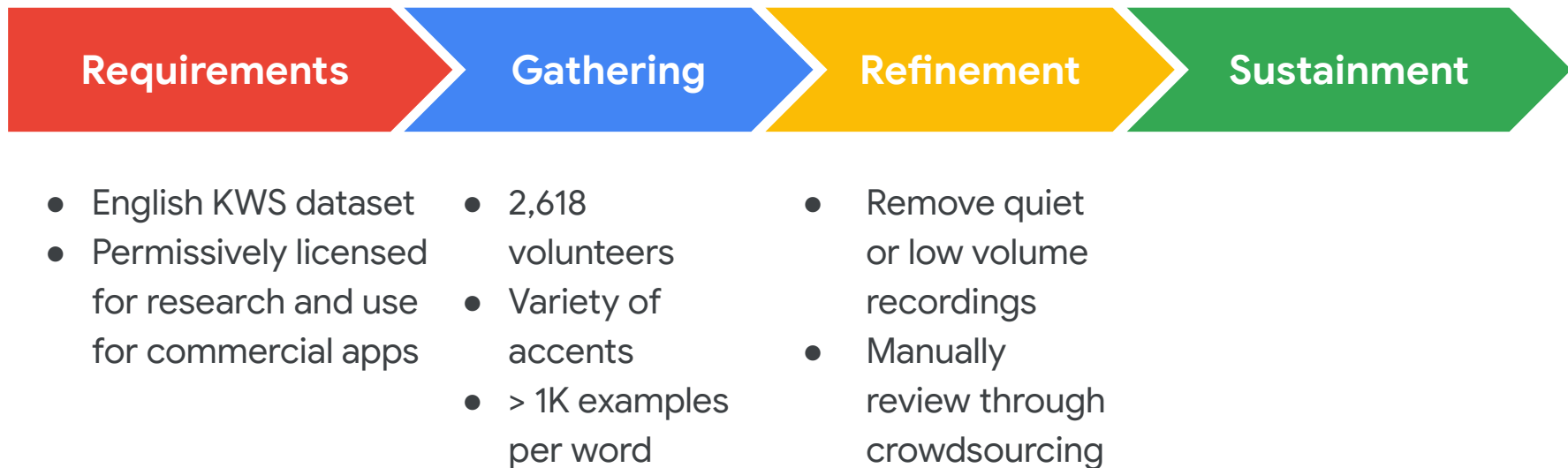
- English KWS dataset
- Permissively licensed for research and use for commercial apps

Data Engineering

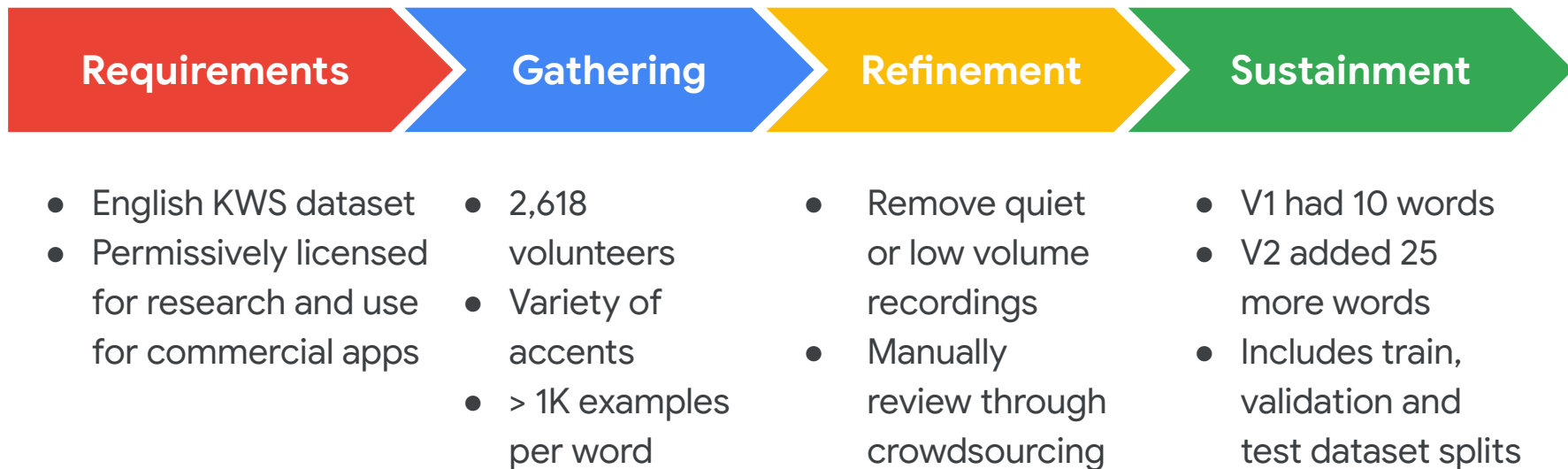


- English KWS dataset
- Permissively licensed for research and use for commercial apps
- 2,618 volunteers
- Variety of accents
- > 1K examples per word

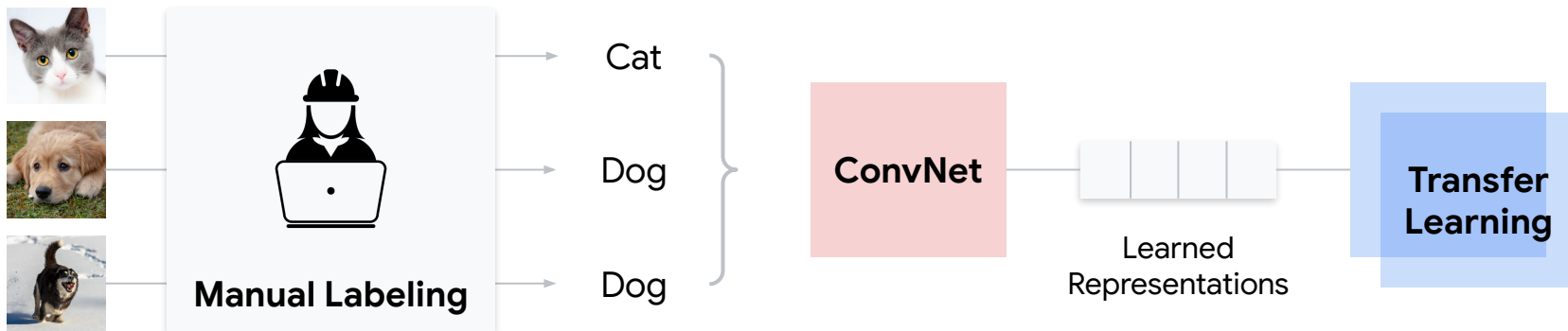
Data Engineering



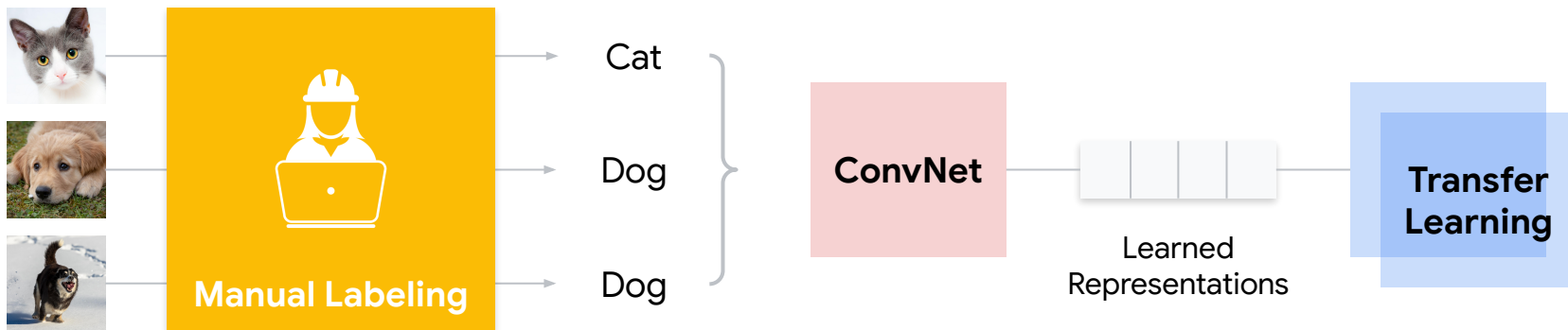
Data Engineering



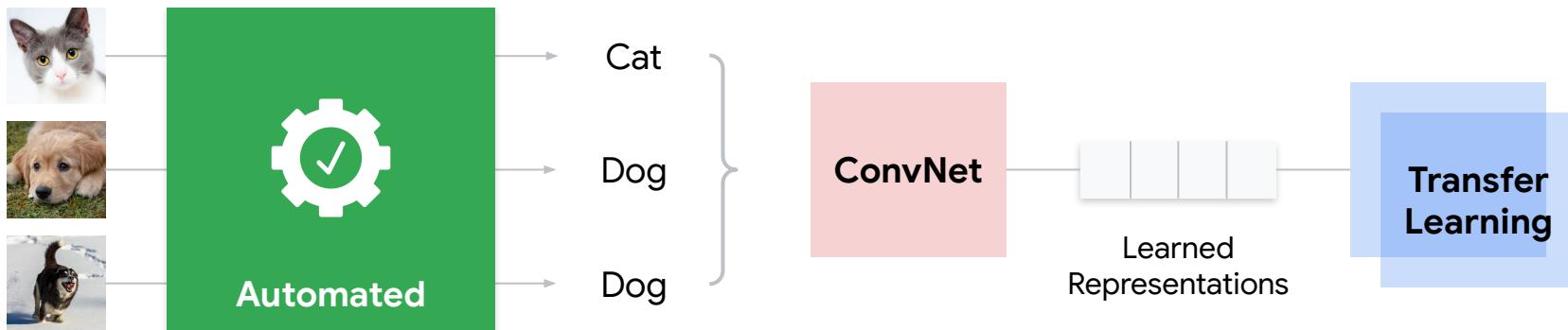
Labeling is Expensive

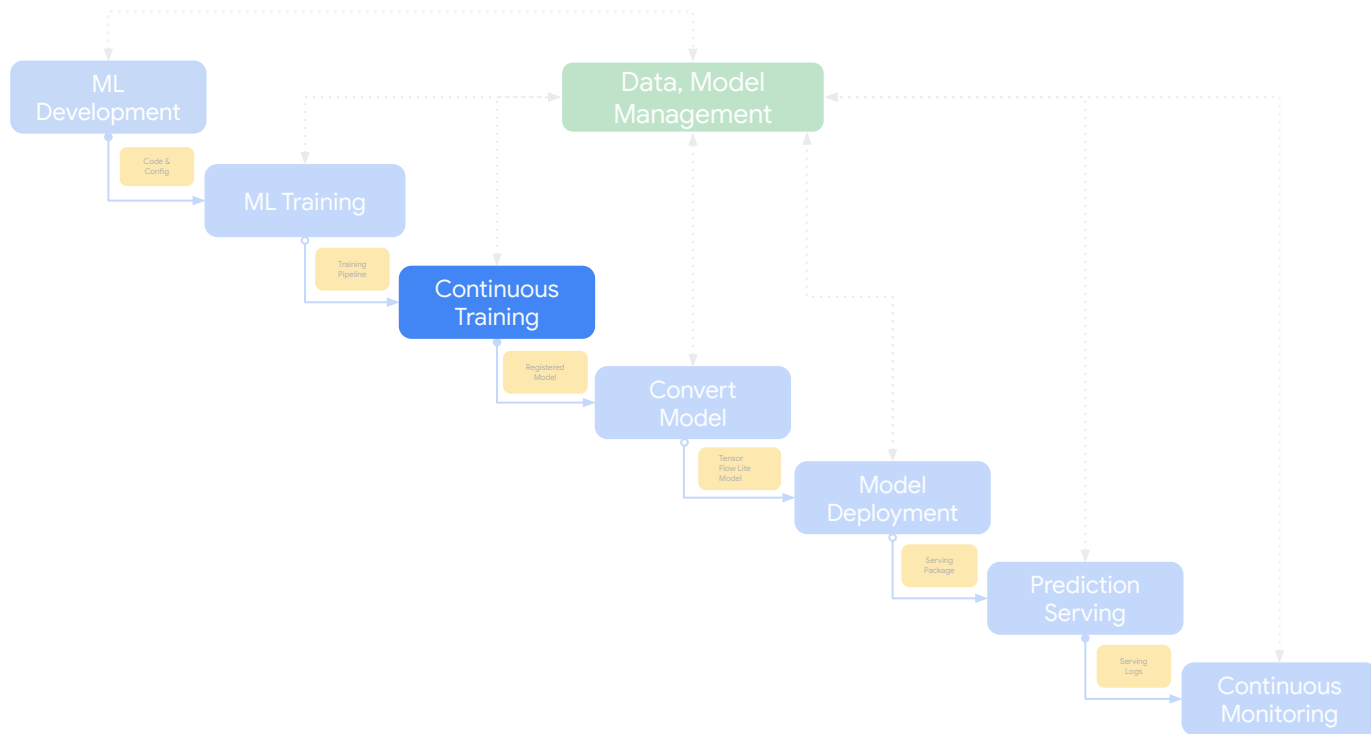


Labeling is Expensive



Labeling is Expensive





Labeling is Expensive

