

0.1 Protobuf

Data for analyzing Protobuf files was collected from 15 different GitHub repositories and is shown in Table 1. In total, there are 8,575 protobuf files which define at least one message with one field.

GitHub Repository	Number of Proto files
googleapis/googleapis	6430
bufbuild/buf	1086
protocolbuffers/protobuf	399
uber/prototool	131
protocolbuffers/protobuf-go	128
google/protobuf.dart	106
samchon/typia	78
colinmarc/hdfs	56
asynkron/protoactor-go	36
streamdal/plumber	35
googleapis/google-cloud-go	25
twitchtv/twirp	23
go-kratos/kratos	17
google/rejoiner	15
cloudwego/kitex	10

Table 1: Overview of data sources for Protobuf evaluation.

0.2 Go + JSON

Data for analyzing JSON encoding from Go was collected for 25 different GitHub repositories and is shown in Table 2. In total, there are 12,954 files across these repositories which use JSON struct tags in top-level struct definitions. All repositories were selected from among the top 1000 GitHub repositories when sorted by number of stars.

GitHub Repository	Number of files with JSON tag
kubernetes/kubernetes	2926
elastic/go-elasticsearch	2490
kubernetes/autoscaler	1473
eksctl-io/eksctl	919
rancher/rancher	805
googleapis/google-api-go-client	634
certimate-go/certimate	392
containers/podman	345
smartcontractkit/chainlink	317
cubefs/cubefs	274
harness/harness	263
gravitational/teleport	253
zitadel/zitadel	250
hashicorp/vault	207
fleetdm/fleet	186
OpenListTeam/OpenList	179
ovh/cds	174
go-gitea/gitea	169
elastic/beats	137
keybase/client	132
caddyserver/caddy	91
tailscale/tailscale	87
cockroachdb/cockroach	75
TencentBlueKing/bk-cmdb	72
rclone/rclone	62

Table 2: Overview of data sources for JSON evaluation.