mozilla

Dati@Mozilla: come affronto il dilemma etico?

Alessio Placitelli - @dexterp37

Linux Day, 23 Ottobre 2021
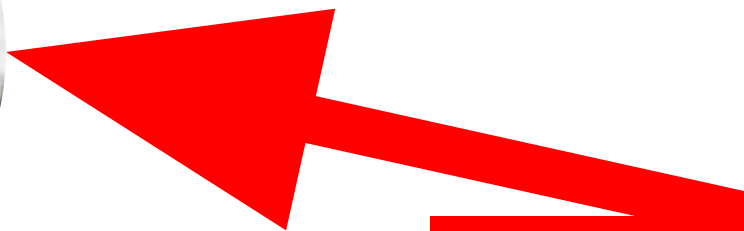
Non un avvocato!

Perché raccogliamo dati?

I nostri principi

https://www.mozilla.org/privacy/principles

1. **Nessuna sorpresa**

2. **Controllo nelle mani degli utenti**

3. **Raccolta dati minimale**

4. **Impostazioni ragionevoli**

5. **Protezione accurata**

The Three Principles

Stay Lean
Decide if all your data collection delivers value.

Build Security
Learn how to protect customer data.

Engage Your Users
Keep customers informed and empowered.

Come implementarli?

Data reviews

https://wiki.mozilla.org/Data_Collection

1. Si crea un bug su Bugzilla o una issue su Github

2. Si compila la data-review request

3. Si notifica un data-steward

La data-review request (1/2)

https://github.com/mozilla/data-review/blob/main/request.md

All questions are mandatory. You must receive review from a data steward peer on your responses to these questions before shipping new data collection.

1. What questions will you answer with this data?

2. Why does Mozilla need to answer these questions? Are there benefits for users? Do we need this information to address product or business requirements? Some example responses:

- Establish baselines or measure changes in product or platform quality or performance.

- Provide information essential for advancing a business objective such as supporting OKRs.

- Determine whether a product or platform change has an effect on user or browser behavior.

3. What alternative methods did you consider to answer these questions? Why were they not sufficient?

La data-review request (2/2)

https://github.com/mozilla/data-review/blob/main/request.md

4. Can current instrumentation answer these questions?

5. List all proposed measurements and indicate the category of data collection for each measurement, using the [Firefox data collection categories](#) found on the Mozilla wiki.

Note that the data steward reviewing your request will characterize your data collection based on the highest (and most sensitive) category.

6. Please provide a link to the documentation for this data collection which describes the ultimate data set in a public, complete, and accurate way.
7. How long will this data be collected? Choose one of the following:
8. What populations will you measure?
9. If this data collection is default on, what is the opt-out mechanism for users?
10. Please provide a general description of how you will analyze this data.
11. Where do you intend to share the results of your analysis?
12. Is there a third-party tool (i.e. not Telemetry) that you are proposing to use for this data collection? If so:

Data categories

https://wiki.mozilla.org/Data_Collection

- Category 1 - technical data

- Category 2 - interaction data

- Category 3 - stored content & communications

- Category 4 - highly sensitive or clearly identifiable personal data

https://github.com/mozilla/glean

Firefox for Android (Fenix)

Firefox Reality

Firefox Lite

Firefox for iOS

mac OS

iOS

Firefox for Desktop

Firefox Developer Productivity (mozregression, mozphab, mach)

PROJECT FOG

Glean registry files

```
15    expires:   2022-02-01
16    application_on_create:
17     type: timing_distribution
18      time_unit: millisecond
19      description: |
20        The duration of `FenixApplication.onCreate` in the main process. This does
21        not measure the duration of migration code (via
22        `MigratingFenixApplication` included in the Beta and Release channels.
23      bugs:
24        - https://github.com/mozilla-mobile/fenix/issues/17969
25      data_reviews:
26        - https://github.com/mozilla-mobile/fenix/pull/17973#issue-572183889
27        - https://github.com/mozilla-mobile/fenix/pull/20623#issue-701630599
28      data_sensitivity:
29        - technical
30      notification_emails:
31        - perf-android-fe@mozilla.com
32        - mcomella@mozilla.com
33      expires: "2022-02-01"
34    app on create to glean init:
```

Specific Metric API

```kotlin
override fun onCreate() {
    // We use start/stop instead of measure so we don't measure outside the main process.
    val completeMethodDurationTimerId = PerfStartup.applicationOnCreate.start() // DO NOT MOVE ANYTHING ABOVE HERE.

    super.onCreate()

    setupInAllProcesses()

    // ... altro codice ...

    setupInMainProcessOnly()

    // DO NOT MOVE ANYTHING BELOW THIS stop CALL.
    PerfStartup.applicationOnCreate.stopAndAccumulate(completeMethodDurationTimerId)
}
```

Come affronto il dilemma etico?

II Glean Dictionary

Nessuna scadenza?

```
    time_unit: millisecond
    description: 'The duration of `FenixApplication.onCreate` in the main
      process. This does

      not measure the duration of migration code (via

      `MigratingFenixApplication` included in the Beta and Release channels.

      '
    bugs:
    - https://github.com/mozilla-mobile/fenix/issues/17969
    data_reviews:
    - https://github.com/mozilla-mobile/fenix/pull/17973#issue-572183889
    - https://github.com/mozilla-mobile/fenix/pull/20623#issue-701630599
    data_sensitivity:
    - technical
    notification_emails:
    - perf-android-fe@mozilla.com
    - mcomella@mozilla.com
```

    Missing required properties: expires

Documentation for this node:
    Describes a single metric.

    See https://mozilla.github.io/glean_parser/metrics-yaml.html
ERROR running glean_parser v2.5.0
```

# deletion-request

apps → fenix 🤖 → **deletion-request**

## deletion-request

`glean-core` 📋

This ping is submitted when a user opts out of sending technical and interaction data to Mozilla. This ping is intended to communicate to the Data Pipeline that the user wishes to have their reported Telemetry data deleted. As such it attempts to send itself at the moment the user opts out of data collection.

## Metadata ⓘ

| Source ⓘ | mozilla/glean/glean-core/pings.yaml#L105 |
|---|---|
| First Added ⓘ | 2019-11-28 12:01:39 |
| Includes Client Identifier ⓘ | Yes |
| Bugs ⓘ | bugzil.la/1587095<br>bugzil.la/1702622 |
| Data reviews ⓘ | bugzil.la/1587095#c6<br>bugzil.la/1702622#c2 |
| Notification Emails ⓘ | glean-team@mozilla.com |

# Periodo di conservazione esplicito

```
44
45    - app_name: focus_android
46      canonical_app_name: Firefox Focus for Android
47      app_description: Firefox Focus on Android. Klar is the sibling application
48      url: https://github.com/mozilla-mobile/focus-android
49      notification_emails:
50        - jalmeida@mozilla.com
51        - tlong@mozilla.com
52      branch: main
53      metrics_files:
54        - app/metrics.yaml
55      ping_files:
56        - app/pings.yaml
57      dependencies:
58        - glean-core
59        - org.mozilla.components:service-glean
60        - org.mozilla.components:lib-crash
61      retention_days: 180
62      channels:
63        - v1_name: firefox-focus-android
64          app_id: org.mozilla.focus
```

"Perché".. !?
Prima o dopo?

Grazie!