



Kanton Zürich

Von Open Data hin zu Open Data Visualization

Gianluca Maccauda - Amt für Gesundheit

Corinna Grobe - Statistisches Amt

- 01** Ausgangslage & Motivation
- 02** Prozess-Beispiele
- 03** Automatisierte Prozesse

Ausgangslage



Open Data



Visualisierungen

Open Code

Fähig & kompetent;
OGD aber nicht
sichtbar (“bekannt”)



Bedürfnis & Ambitionen;
fehlendes Know-how



(Open) Data User

Motivation

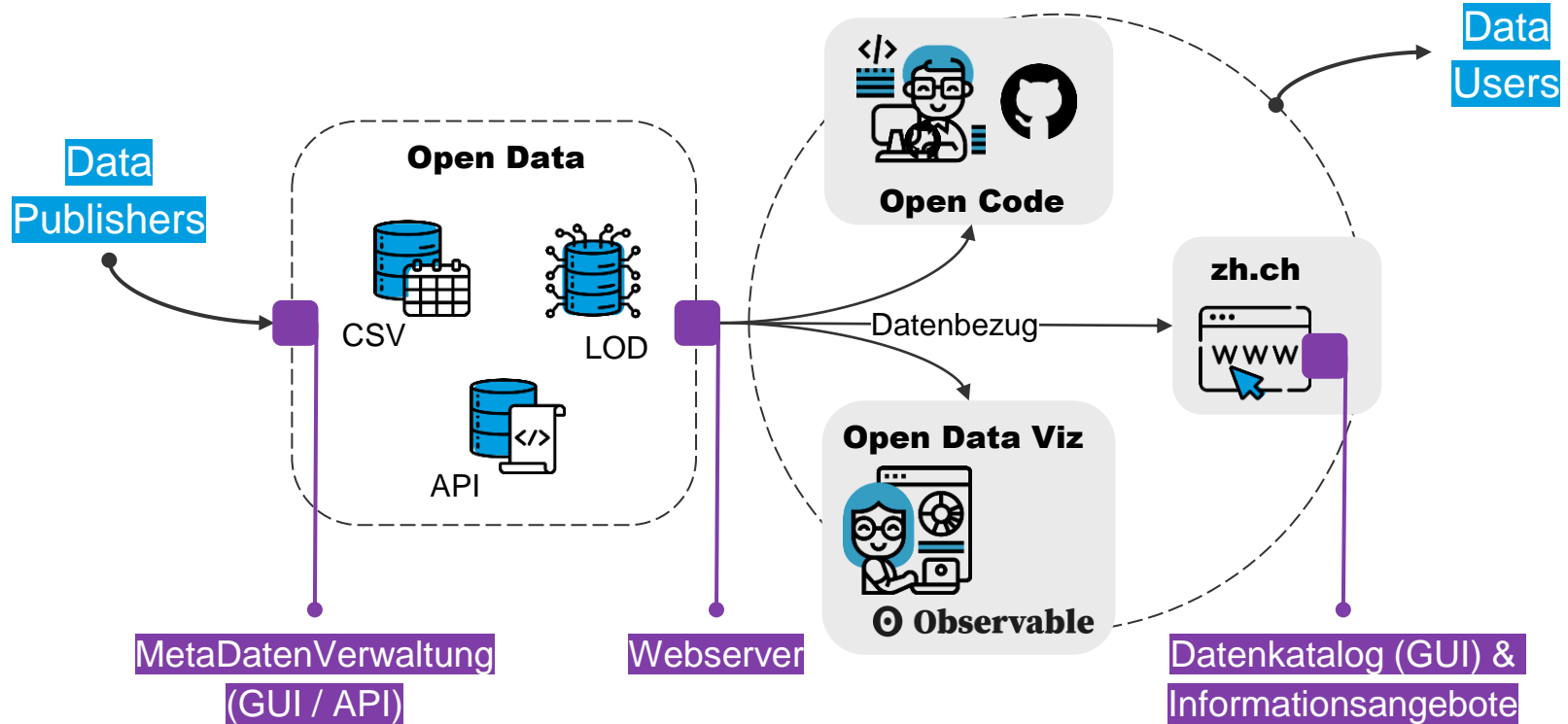
- Die Zahl der OGD-Datensätze wächst und damit auch die Nachfrage nach Daten und datengestützten Informationsprodukten, wie z. B. Datenvisualisierungen.
- Notwendigkeit eines durchgängigen Prozesses von OGD bis zum datengestützten Informationsprodukt.
- Aufbau von Prozessen, mit denen wir bei geringem Aufwand nachhaltig und effizient arbeiten können. → Standardisieren & Automatisieren

→ **«Open Data Ökosystem»**


«Open Data Ökosystem»

« Ein Datenökosystem ist das prägende, **ganzheitliche Umfeld**, in dem verschiedene **Akteure zusammenkommen**, um **Daten zu produzieren, anzubieten, zu finden und zu “konsumieren”** (d. h. nachzunutzen, zu verarbeiten, anzureichern, zu archivieren, zu publizieren, Entscheidungen darauf zu fällen etc.). Die Einflüsse des Datenökosystems wirken **in alle Phasen der Datenlebenszyklen** hinein, es schafft die entsprechenden Rahmen-, Netzwerk- und regulativen Bedingungen für die (Zusammen-)Arbeit mit Daten bzw. stellt diese konkret dar.»


«Open Data Ökosystem»



Open Data as Main Ingredient



« Die Philosophie "open data as main ingredient" hat uns geholfen, unsere Datenveröffentlichungsprozesse effizienter zu gestalten. Sobald neue Datenpunkte verfügbar waren, konnten wir uns auf die Aktualisierung der offenen Daten konzentrieren - und nur darauf. Jede Visualisierung oder Anwendung, die auf diesen Daten basiert, würde automatisch aktualisiert werden. Selbst für Datensätze, die einmal im Jahr aktualisiert werden, hat sich dies als praktisch erwiesen.»



GitHub als OGD-Hub

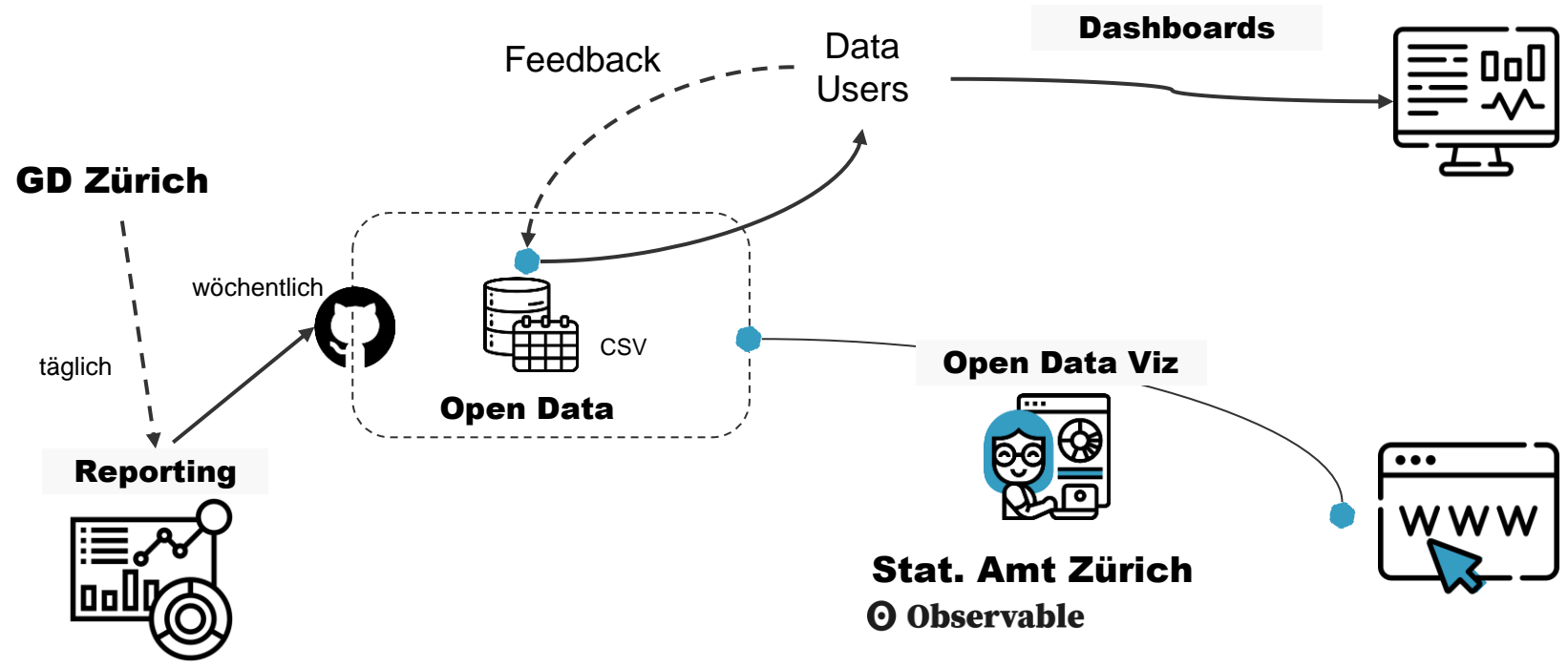
Covid-19 OGD als Beispiel

- Pandemie als OGD Beschleuniger:
 - Grosse Verunsicherung, Öffentlichkeit will informiert sein
 - OGD: Dashboards (Medien, private Dashboards, Kanton ZH)
- Hochfrequente Daten: Tägliche Aktualisierungen, viele Quellen, Eingabefehler
 - Keine vollständige Automatisierung möglich
- Endprodukt: Webseite zur Situation von Covid-19 im Kanton Zürich
 - vollständig aus offenen Daten gefüttert

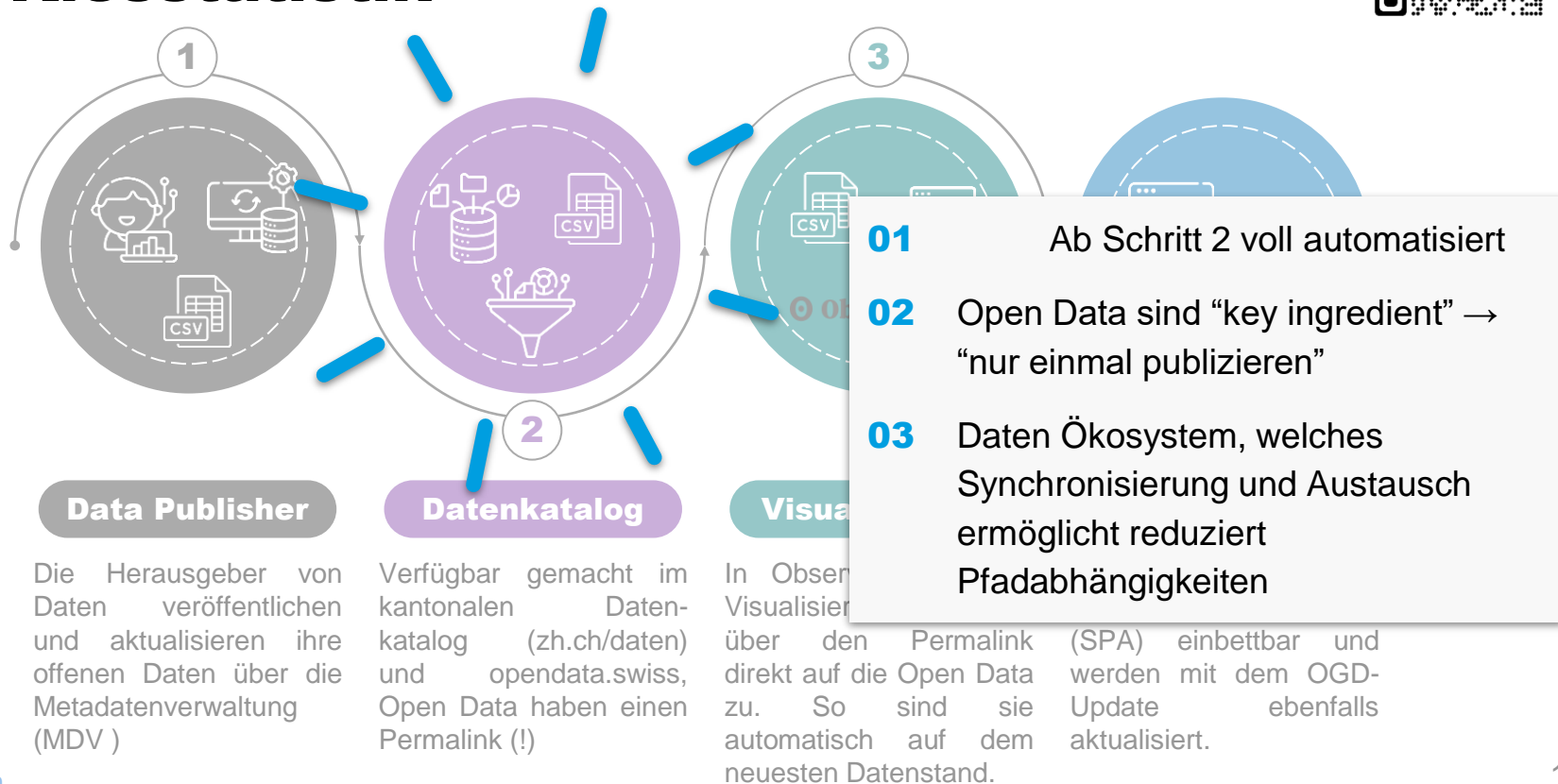




Covid-19 OGD als Beispiel



Beispiel 1: Linearer Prozess Kiesstatistik



Beispiel 2: Loop-Prozess Kriminalstatistik



5

Herausforderungen

- 01** Code offen zur Verfügung stellen?
Was und wie?
- 02** Datenaufbereitung z.T. auch unter
Einbezug nicht-öffentlicher Daten
- 03** DB-Abfragen nicht public machen
(Sicherheit !)
- 04** Wie den Aktualisierungsprozess
automatisieren?

1



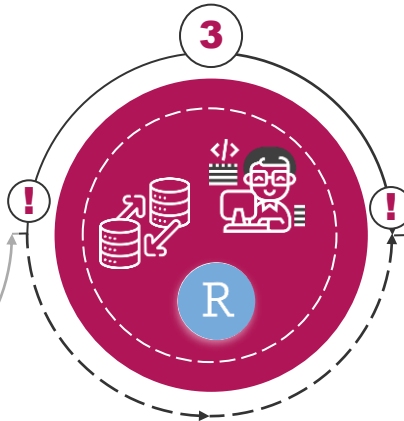
Data Publisher

2



Datenkatalog

3



Datenbearbeitung

Datenaufbereitung oder -
anreicherung, u.a. aus
internen Datenbank.

Prozesse mit Pipeline-Tools automatisieren

Status-Quo

- Data Management: Mehrere Prozesse etabliert (Metadatenverwaltung, Gemeindeporträt, Open Data)
- Andere Abteilungen: Mehrere Prozesse in persönlicher (!) Crontab auf dem R-Server

Probleme

- Keine gesamtheitliche Übersicht über die eingerichteten Jobs im Amt
- Root oder persönliche Rechte notwendig
- Zunehmende Bedürfnisse an Jobs

Lösung

- Einführung eines (Data)-Pipeline-Tools wie Airflow oder Jenkins
- Entscheid für Jenkins, da bereits Know-How vorhanden

Jenkins

- Schöne Übersicht über Status der einzelnen Jobs
- Jeder Arbeitsschritt (Stage) kann einzeln verfolgt werden
- Notifications (Fail/Success) per Mail oder Slack
- Eigene Nutzer/Gruppen- und Passwortverwaltung
- Jobs werden in relativ einfacher Scriptsprache verfasst
- Git-Integration
- Zugriff auf R-Server, Data Management Server, DB's, Webserver, etc.
- Linux-Skills (Shell) notwendig



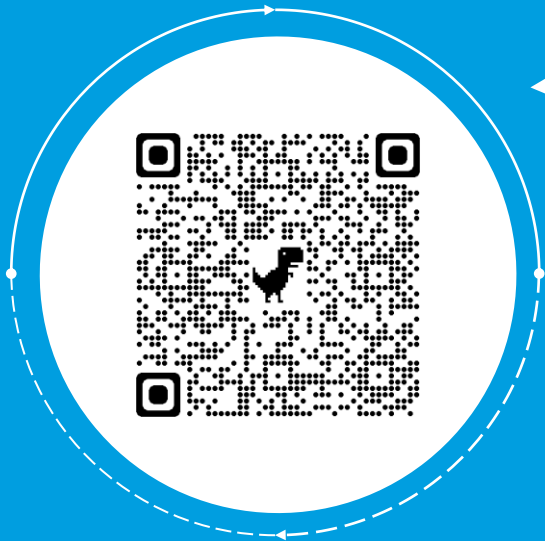
Obwohl wir vieles richtig machen, können wir immer noch besser werden



- Datenveröffentlichungsprozesse effizienter - nur OGD aktualisieren
- Code für Visualisierung bildet auch Template. Dient damit als Lernressource und kann von anderen Direktionen / Nutzenden verwendet und angepasst werden.
- Systematischer Ansatz mit kontinuierlicher Weiterentwicklung - mehr Automatisierung



- Konsequenz Lizenz hinzufügen
- Noch nicht alle Komponenten sind OGD, z.B. Kartengrundlagen
- Code der Visualisierung wird noch nicht an der Data Viz oder OGD verlinkt. (Wer sucht der findet)
- Einheitliche Figures? Geteiltes ggplot2 Theme oder Design System
- Einbindung im Layout von Kanton verbesserungsfähig



Zu den Slides

Wir freuen uns über Feedback, Fragen und Anregungen.

Gianluca Macauda

gianluca.macauda@gd.zh.ch

043 259 52 43

Corinna Grobe

corinna.grobe@statistik.ji.zh.ch

043 259 75 07