health RI

enabling data driven health

Once data are FAIR,

machines and people can find and use them

Projectleidersbijeenkomst Infectieziekten bestrijding & Antimicrobiële resistentie, ZonMw "Over het FAIRificeren van data in de IZB & AMR-projecten van ZonMw"

June 22, 2022 Jeroen Belien

FAIR Hourglass

Raw Data
[freedom to operate]

FAIRification

FAIR Layer 1: Creation of data & capture of data-reuse consent choices made by appropriate stakeholders

FAIR Layer 2: ETL and 'harmonization' processes for data & consent targeting domain-relevant machine-readable vocabularies & schema

Data that are FAIR-Ready
[a minimal standard that is FAIR, open, and technology-independent]

FAIR Center: Storage of data, metadata, and consent as machine-actionable units of information (distributed)

FAIR Digital Object Framework

FAIR Orchestration

FAIR Layer 3: Exposure of FAIR metadata & gateways for ontology-based access control

FAIR Layer 4: Automated FAIR services for indexing, search, semantic resolution, data integration, object routing & workflow execution

Data Integration
Data Sharing/Visiting
Data Analytics
[freedom to operate]

FAIR Layer 5: High-level data use applications such as data landscape surveys, data exchange, distributed learning and data analytics

health RI
enabling data driven health

Road (so far) to machine actionable (meta)data
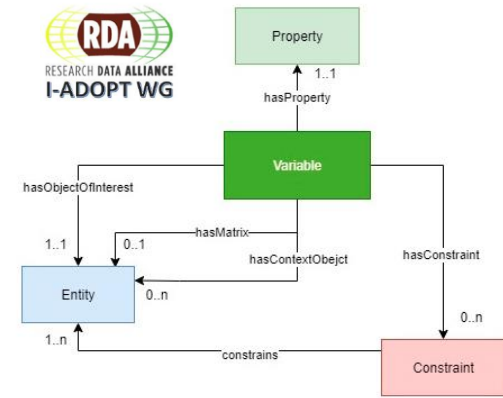
From Community to Data Portal
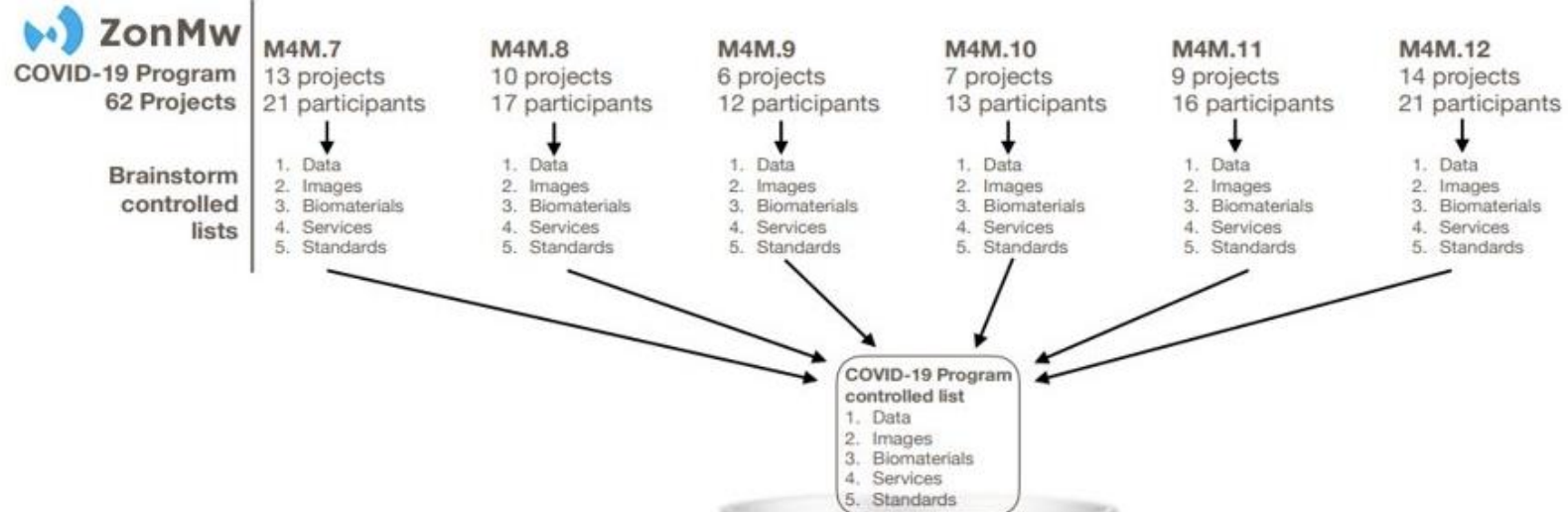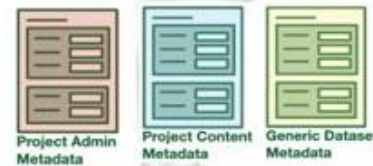
Data ETL workflow from an institute/collaboration to a portal

ZonMw metadata portal: https://covid19initiatives.health-ri.nl/p/ProjectOverview
(machine readable metadata → easier → human readable portal)

semantic faceted search using the provided metadata

and contact data custodian and request (getting access to) data

NFU COVID-NL metadata portal: https://covid19initiatives.health-ri.nl/p/Dashboard
(also built on machine readable metadata from CEDAR)

Search and request
(getting access to) data

So why do we need machine readable data?

- Lessons learned so far:
  - machine readable metadata → easier → human readable portal
  - study metadata already **findable** from *the start* of projects while new data still is being generated
  - metadata is **interoperable**
    - at the same portal
    - metadata no longer is restricted to "that one portal"
      - based on standards, like DCAT, that facilitates interoperability between data catalogues/repositories published on the Web
      - E.g. NFU COVID data -> https://www.covid19dataportal.org/
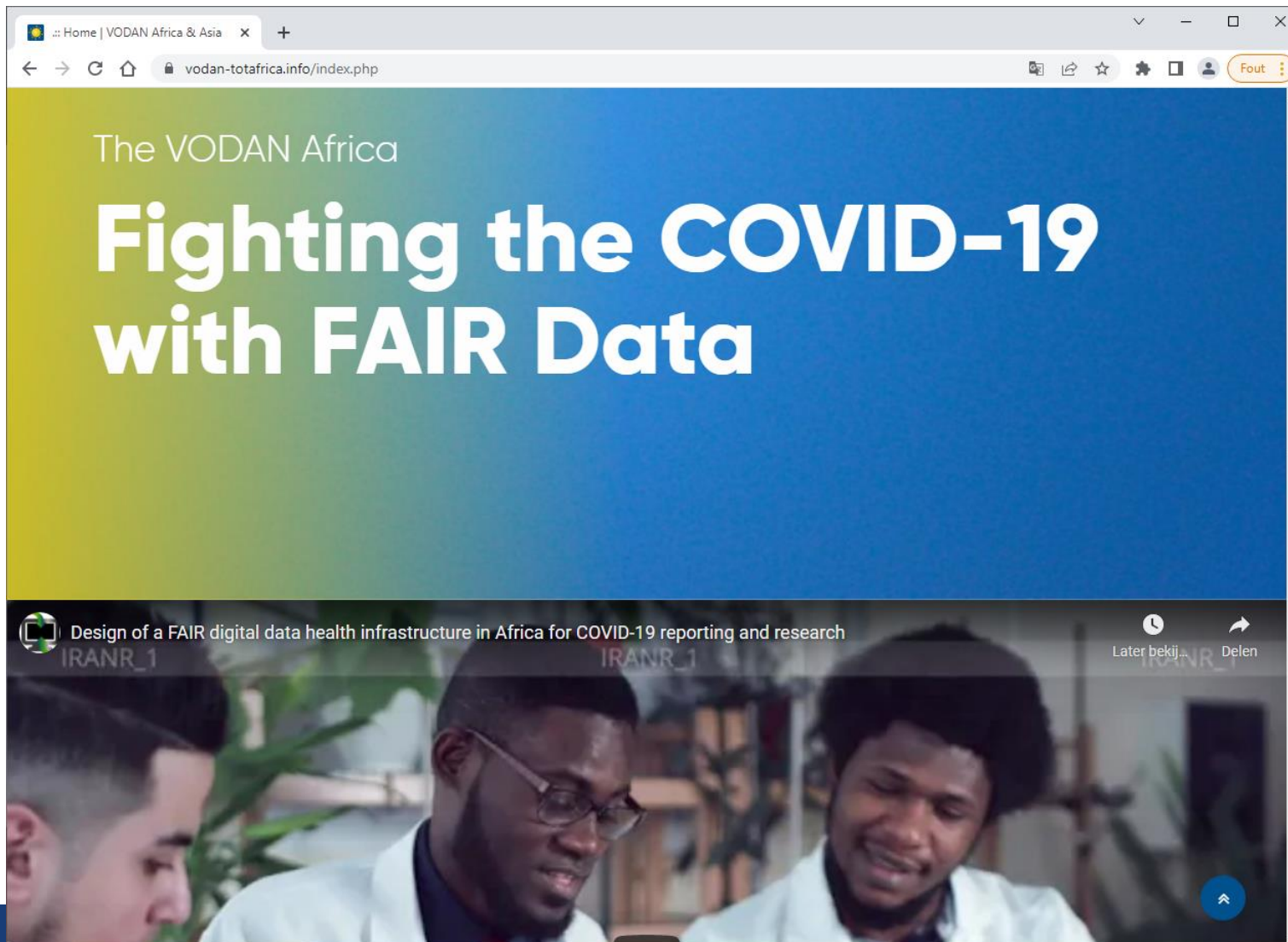 - **discoverability** of metadata increased

Other examples

- Open data:
  corona dashboard

- Rioolwater
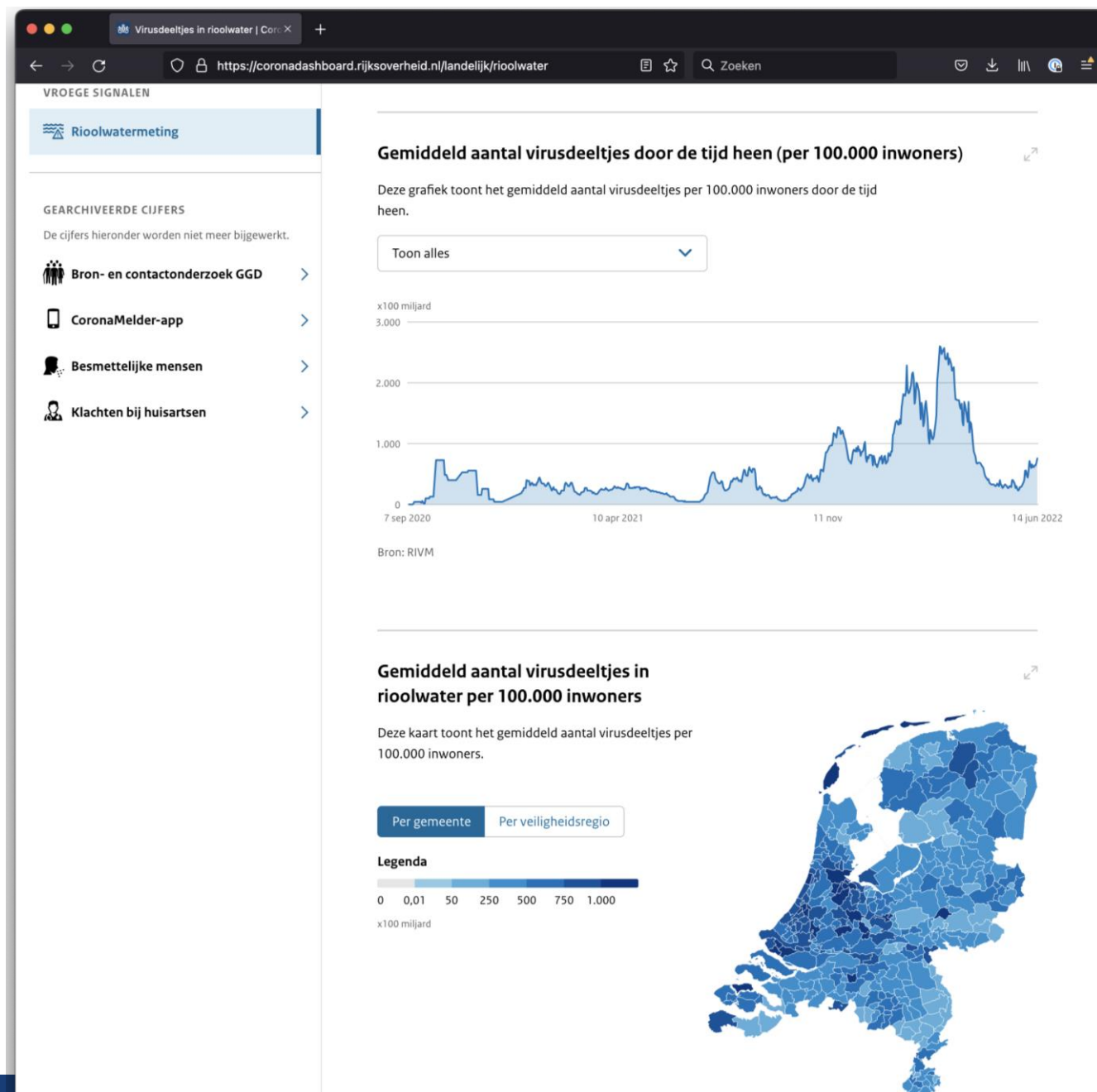
FAIR orchestration: FAIR Data Points etc. by Mirjam van Reisen see what's FAIRly possible now in Africa thanks to VODAN Africa