

REACH



Datavillage



VRT SANDBOX

rtbf.be

umec

istart

CO.STATION

io.e
Internet of Energy



FINTECH
BELGIUM



European
Commission
Horizon 2020
European Union funding
for Research & Innovation



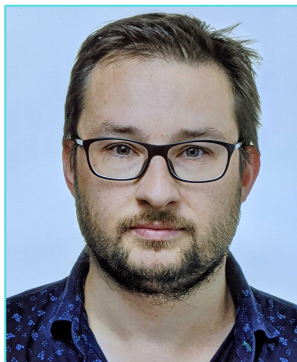
Wallonie
service public
SPW



Collibra

The team

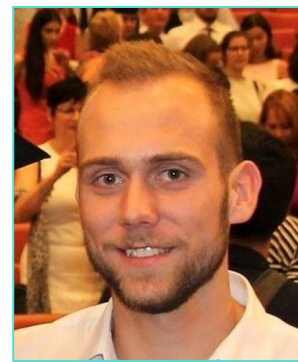
Business data value
Data user experience
Privacy & user control
Personal data regulations



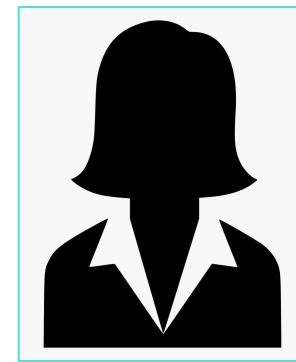
Philippe Duchesne
Technical lead



Loic Quertenmont
Lead data scientist



Justin Sirjacques
Full stack dev



Data engineer



Frederic Lebeau
*Product & Tech
Co-Founder*



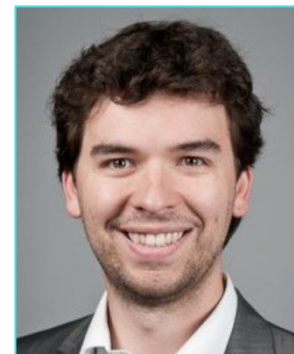
Quentin Felice
*Business & Growth
Co-Founder*



Laurens Van Cauwelaert
UX & Front-end dev



Tijs Gommeren
Front-end dev



Alexandre Cassart
*Consultant
Lawyer & CIPP/e*

Consumers curated life

76% of users

rtbf
duvio

vrt

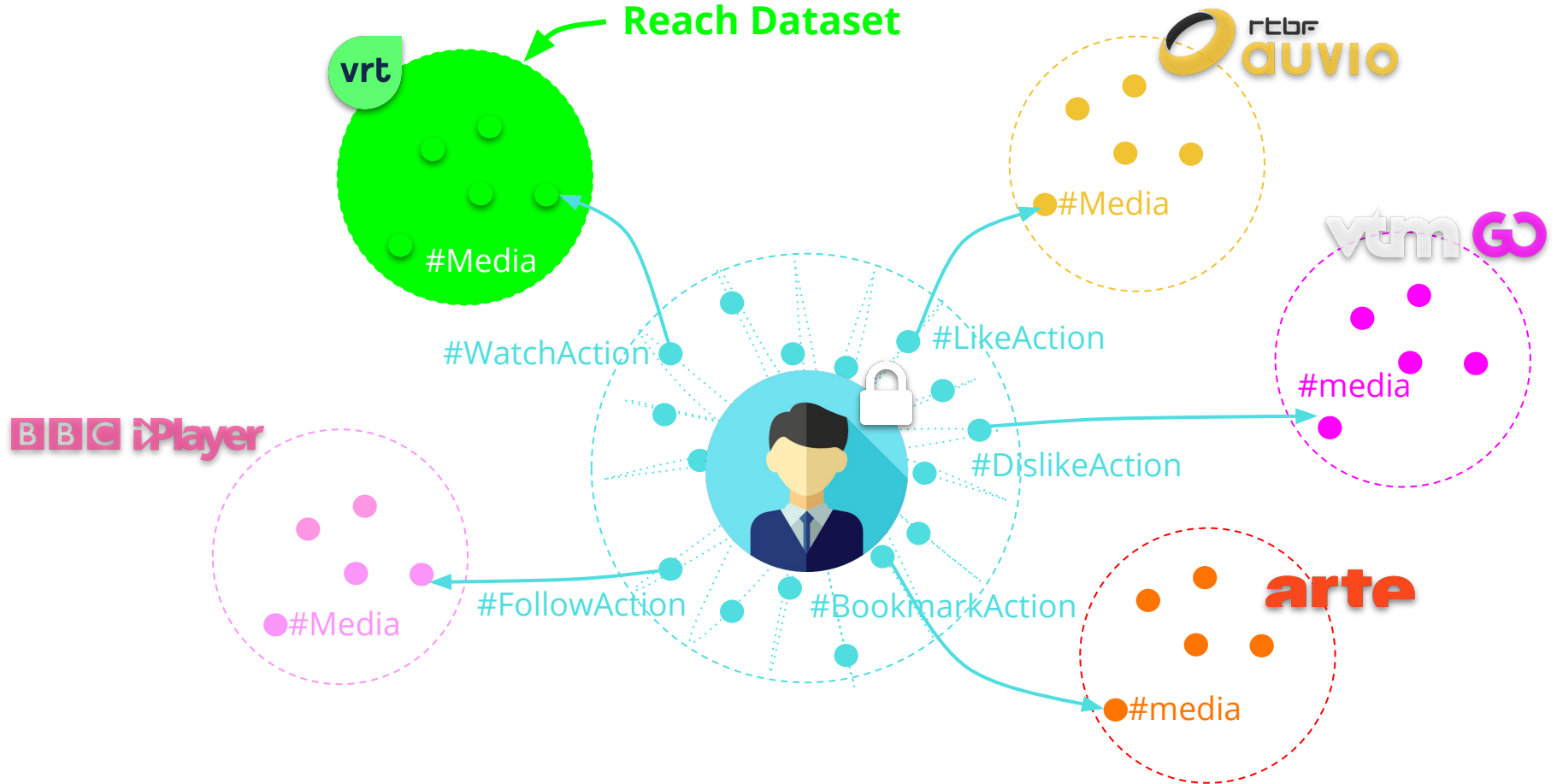
BBC iPlayer

vtm GO

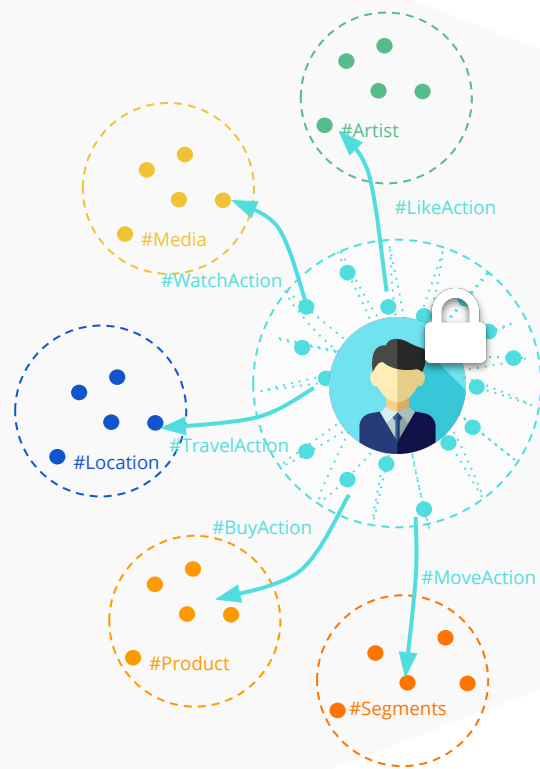
arte

What they miss !

Link the data on consumer's side



Ask users questions through their data !



Privacy



Consent

Question

Discover



Broadcaster 1

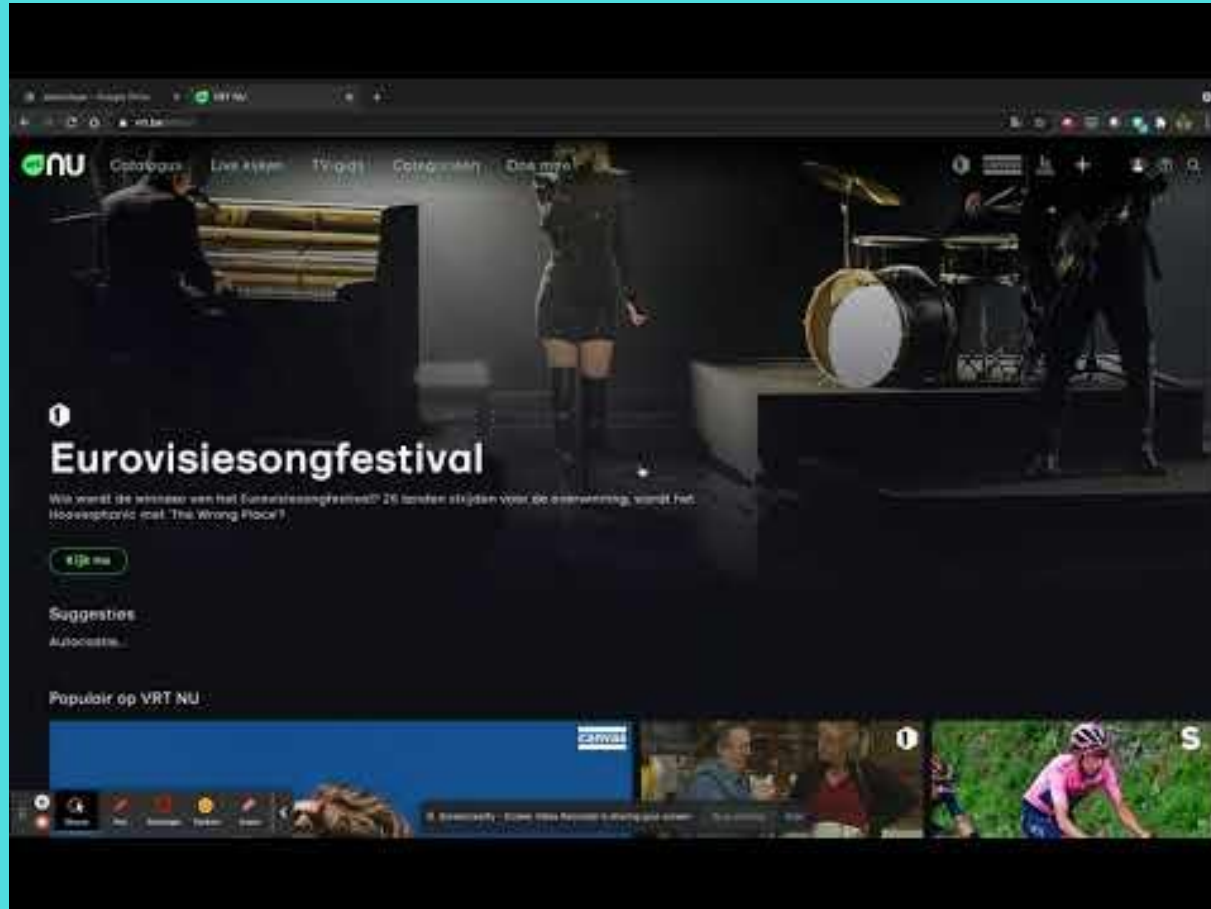


Broadcaster 2

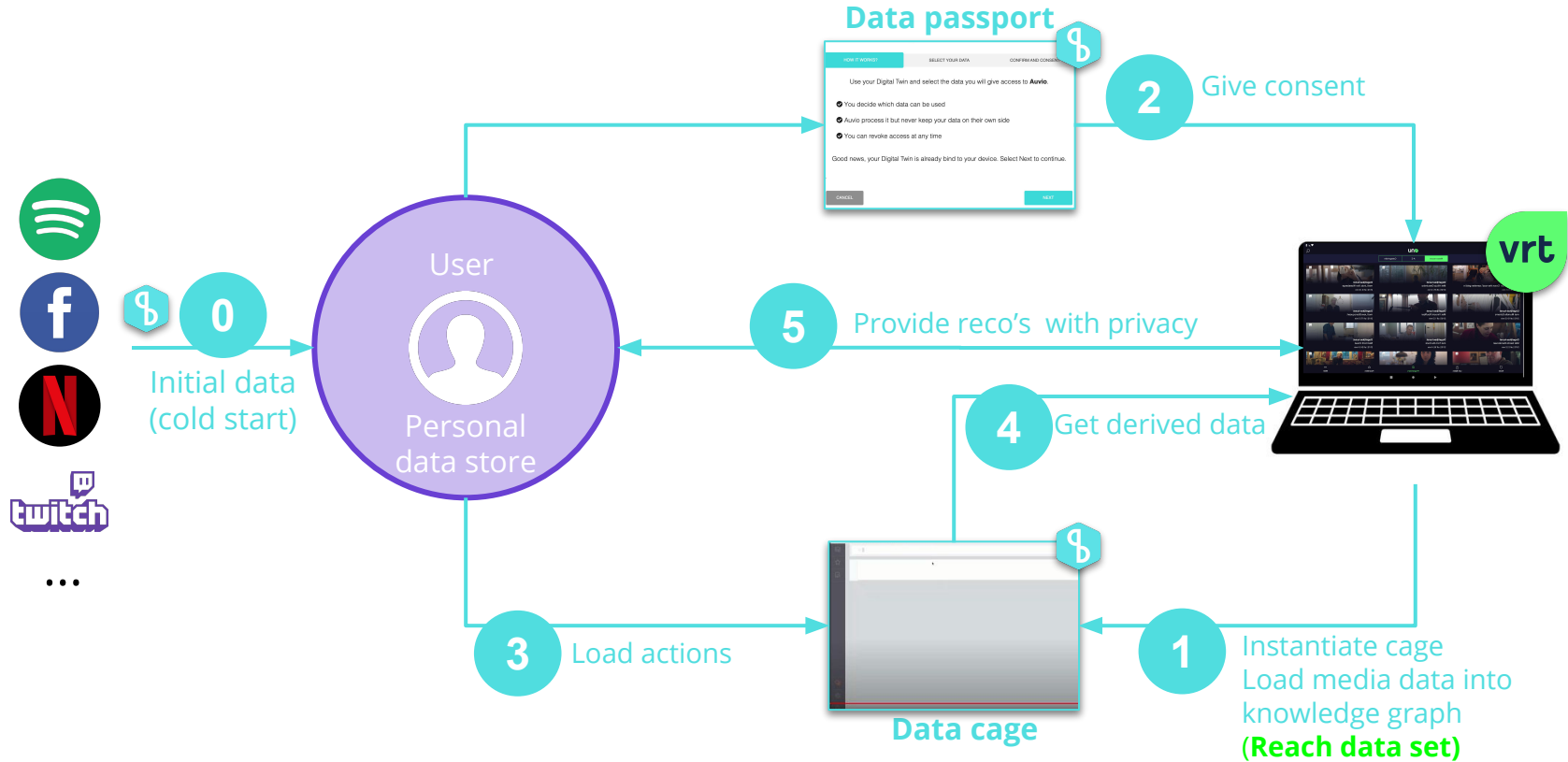


Broadcaster n

The product



How it works





Datavillage technologies

User centric recommendation flow

Data Sources for the POC:

- **Media metadata to recommend**
- *User previous actions on FB, Spotify, Auvio*
- *Knowledge Graph enrichment*
Mick Jagger → Rolling Stones → Rock Music
- *NLP enrichment*
Named Entity Recognition, Bag of Word, TF-IDF

Reach Dataset



Recommendations for a specific user

User

User Actions

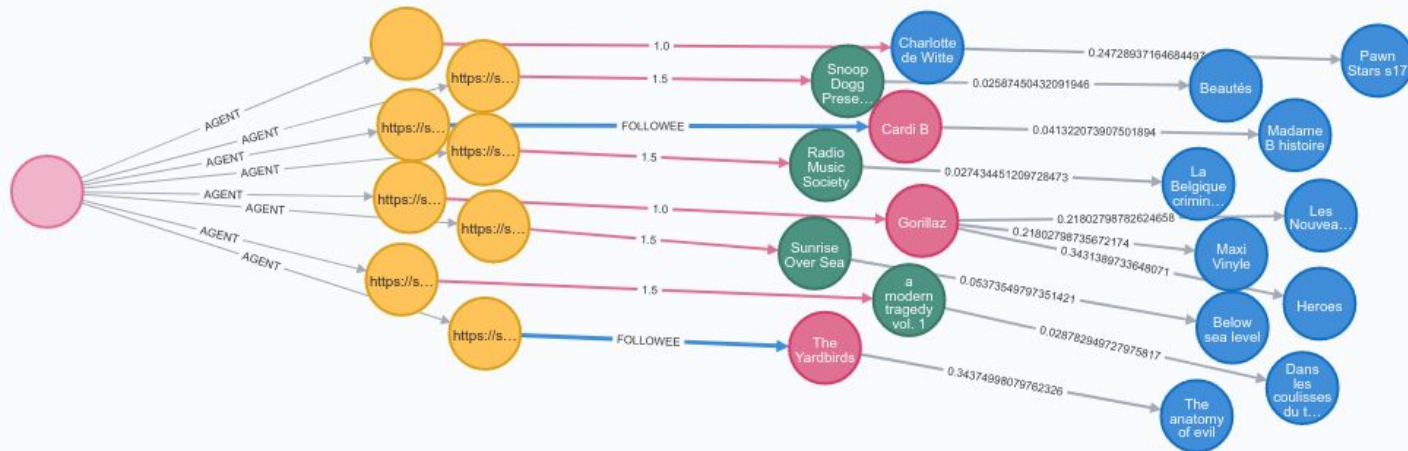
Media Touched

Media Recommendations

Look at previous user interactions with media
FB Likes, Spotify bookmarks, Watch on VRT, Auvio, etc.

Use media touched by the user to find similar content in the graph

Combining similarity measure, user action weight and number of paths to a similar media

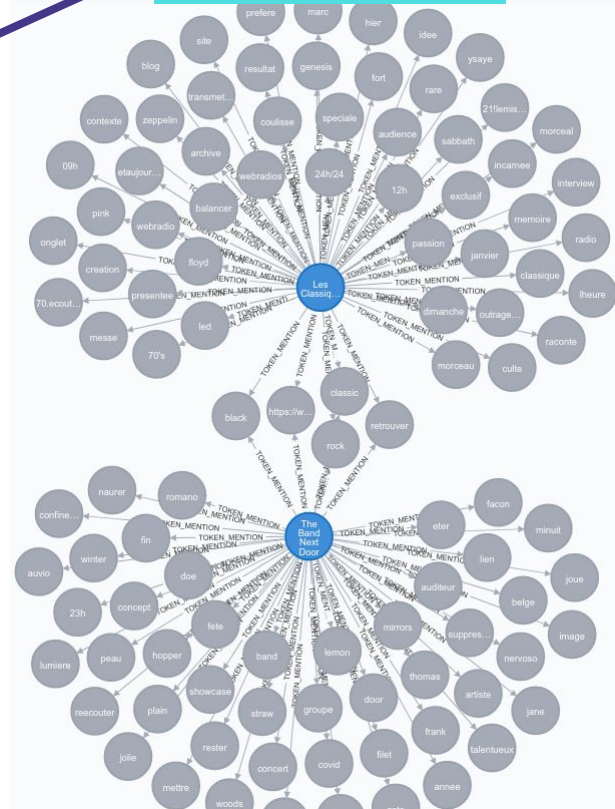


Media Similarity Measure from NLP

Shared Named Entity



Shared Words



TF-IDF

Term Frequency
Inverse Document Frequency
to build vector representing a
single media

Cosine Similarity

between two media
value close to 1 if media share
all their NER/Word content
and close to 0 if they have no
common NER/Words

Media Similarity Measure from Knowledge Graph Concepts

Enrichment

Enrich the knowledge graph from external data sources in order to find connection between nodes



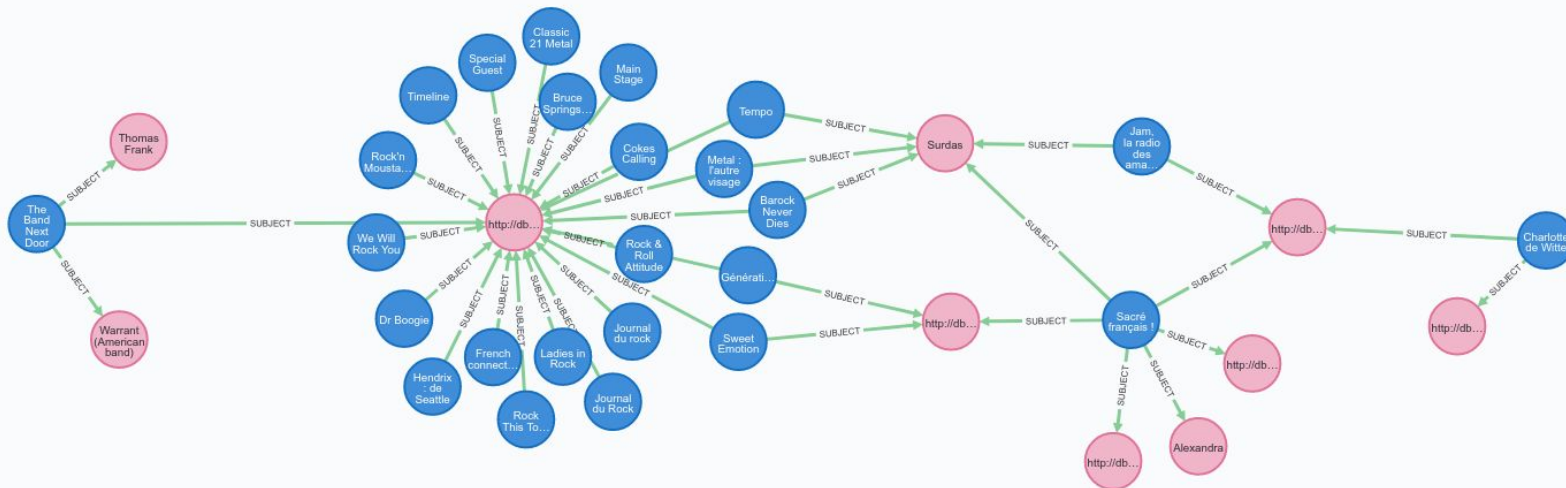
Shortest Path

Find the shortest path
between pairs of nodes
(distance depends on the
quality of the connections)



Similarity

between media is computed
from the distance among
them



Overall Similarity Between Two Medias

Named Entity Similarity

Concept Similarity

Word Similarity

Weighted Sum

Overall Similarity

Best Connections

For every media, we precompute the N most similar media connected to it.



What influences recommendation scoring

User

User Actions

Media Touched

Media Recommendations



- **User Action**

- **Date:** recent is better
- **Type:** bookmark>follow>like
- **Positiveness:** Like vs Dislike
- **Platform:** action on audio are underweighted to avoid bubble effects

- **Media touched by an action**

- **Content:** description, keywords, album vs track vs music band
- **Frequency:** we try to use each touched media only once for each user recommendation → increase diversity
- **Similarity/Connectivity:** with Audio content is used to find recommendation candidates

- **Recommended Media**

- **Publish Date:** recent is better
- **Number of actions / medias:** suggesting this recommendations
- **Recommendation Likelihood:** some media are easily recommended because they have high connectivity/similarity. Those are underweighted to avoid bubble effect.
- **Watched media:** are discarded
- **Expired media:** are discarded

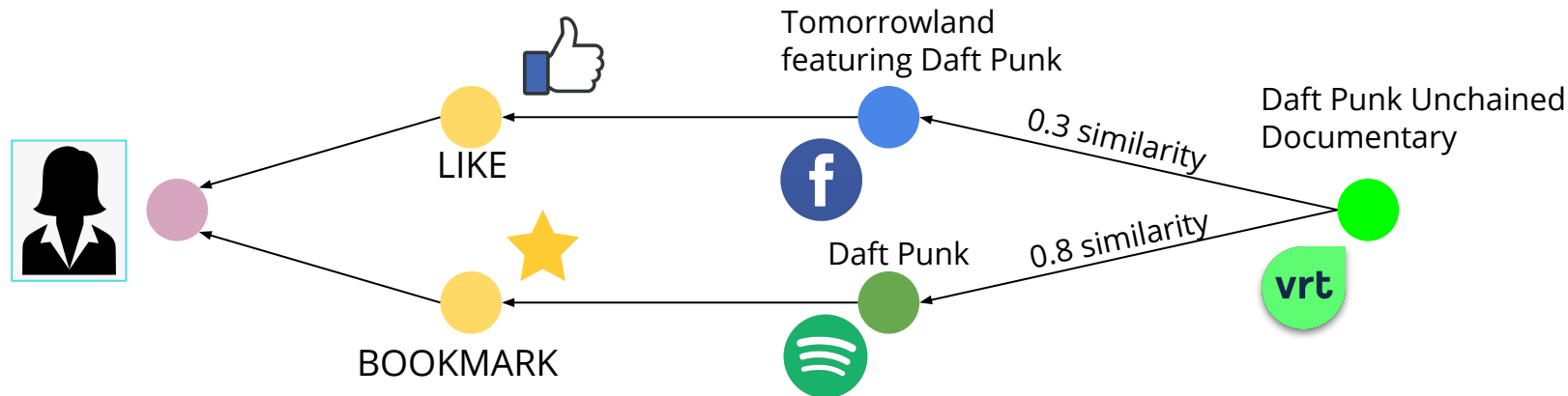
Explainability of the recommendations

Reverse Navigation

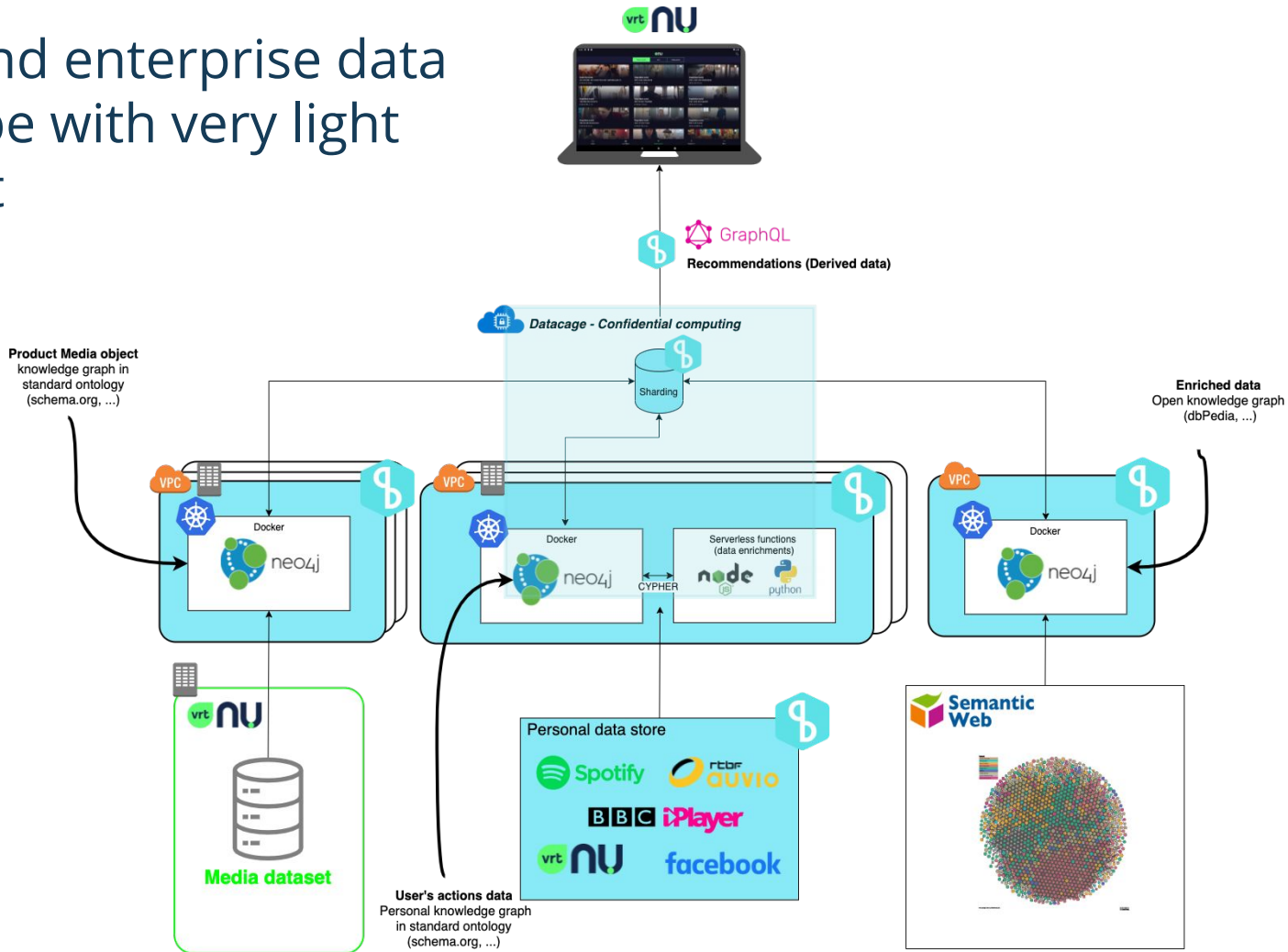
The knowledge graph can be navigated in opposite direction in order to explain why a particular media was recommended to a user.

Actions \leftarrow \rightarrow Recommendations

We can provide to the end-user which actions / data providers are determinant for a specific recommendation



We extend enterprise data landscape with very light footprint



Thank you

www.datavillage.me

  @datavillage

