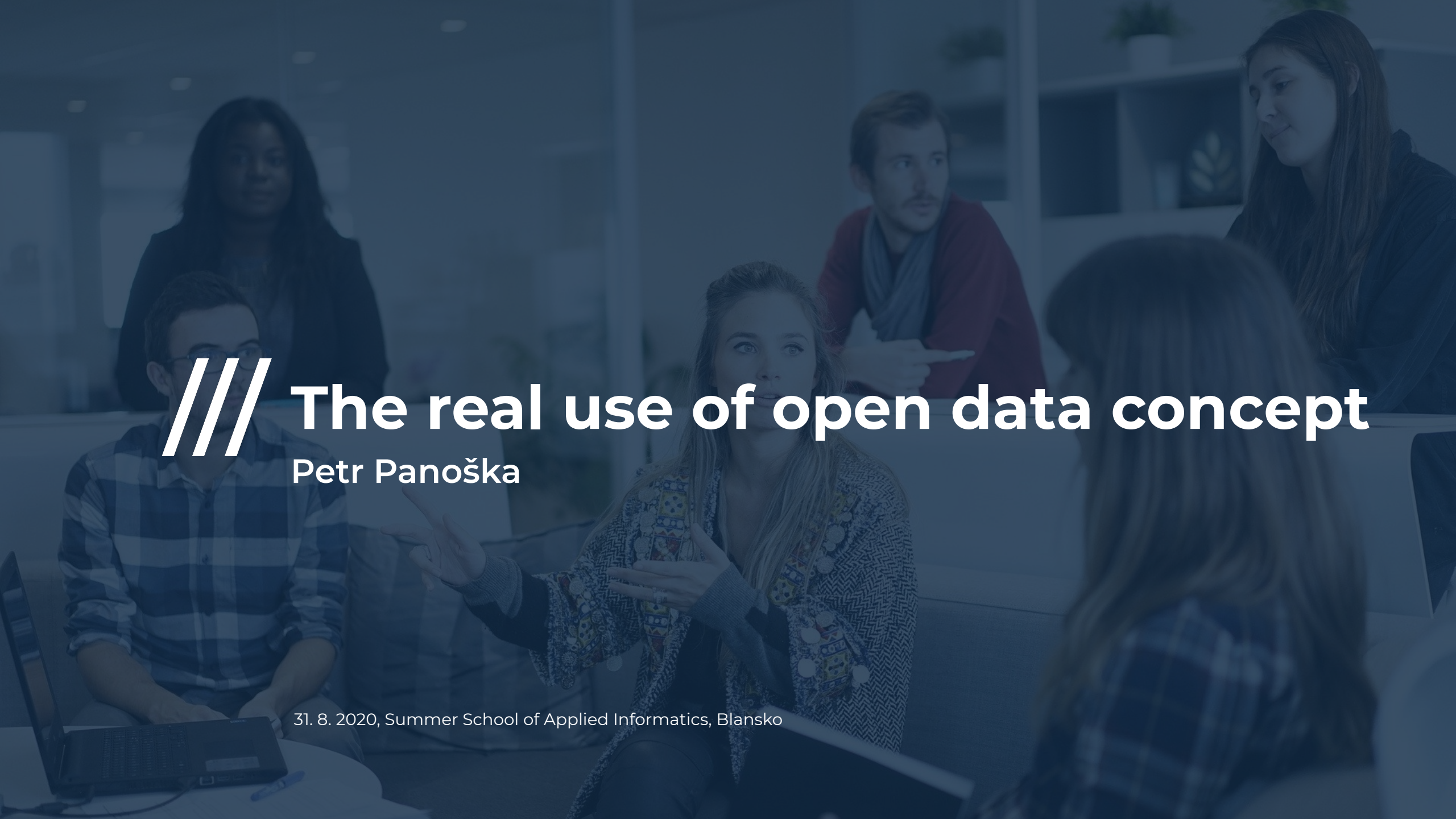




/// The real use of open data concept

Petr Panoška

31. 8. 2020, Summer School of Applied Informatics, Blansko

Intro

- petr.panoska@uzis.cz
- last 3 years in Webstudio as web developer
- mostly oriented to backend (Symfony + Doctrine + Pgsqll)
- occasionally frontend (Vue, React)

Agenda

- The big picture
- Publishing process
- Technical solution and use cases



Open data??
Let me ask you...

/// The big picture

- National catalogue of open data
 - <https://data.gov.cz/>

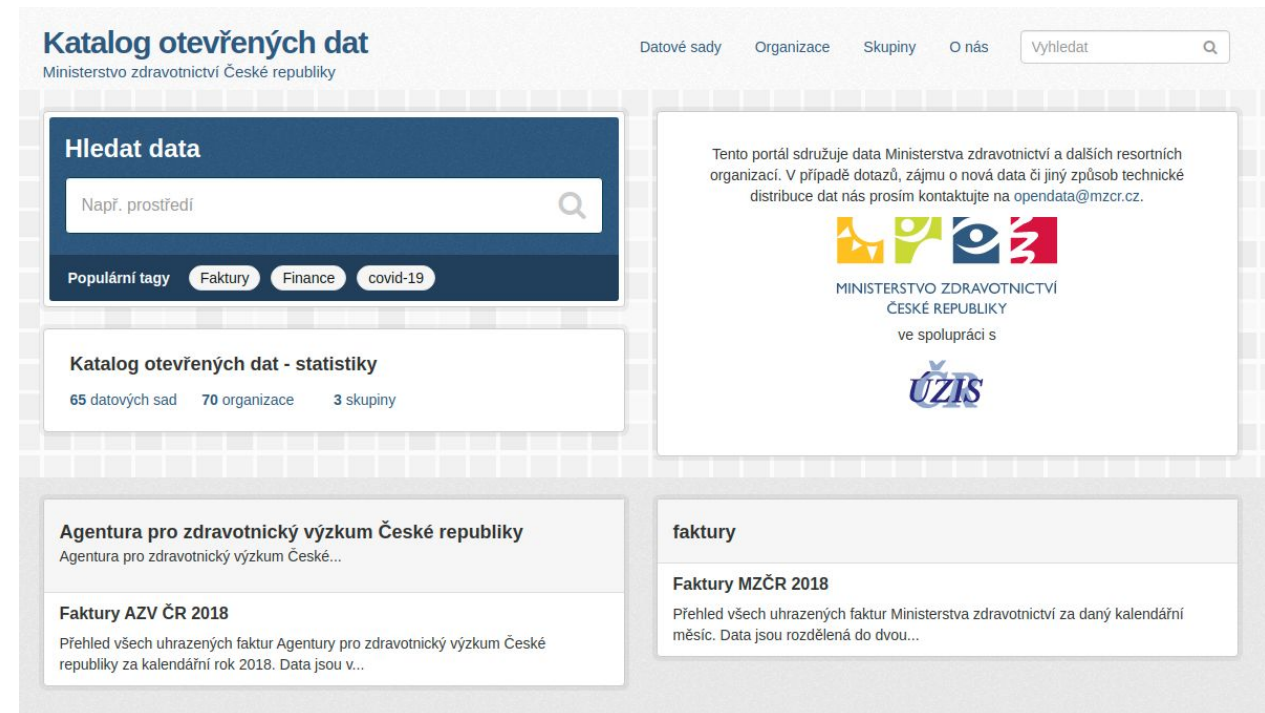
The screenshot shows the homepage of the National Catalogue of Open Data (data.gov.cz). The header includes a logo 'OTEVŘENÁ DATA' and navigation links: 'Články', 'Novinky', 'Národní katalog otevřených dat', and 'Další informace'. The main banner features the text 'Portál otevřených dat' and 'Vše, co potřebujete vědět o otevřených datech v České republice'. A large number '135 121' is displayed next to a map of the Czech Republic, with the text 'datových sad' below it. A search bar with the placeholder 'Název datové sady' and a 'Hledat' button is also visible. Below the banner, there is a section titled 'Novinky ze světa otevřených dat' with three news items:

- Konference Otevřená data 2020 bude 23. 10. 2020**
26. 8. 2020
V pátek 23. října 2020 v Národní technické knihovně v Praze se uskuteční konference Otevřená data 2020! Poznačte si do kalendáře!
- Výroční zpráva o stavu otevřených dat v České republice za rok 2019**
14. 7. 2020
Vláda v pondělí 13. července projednala Výroční zprávu o stavu otevřených dat za rok 2019.
- Podání návrhů projektů využití otevřených dat veřejné správy EU**
25. 6. 2020
Od 30. června se otevírá výzva CEF podání návrhů projektů využití otevřených dat veřejné správy EU.

A button labeled 'Více informací »' is located at the bottom right of the news section.

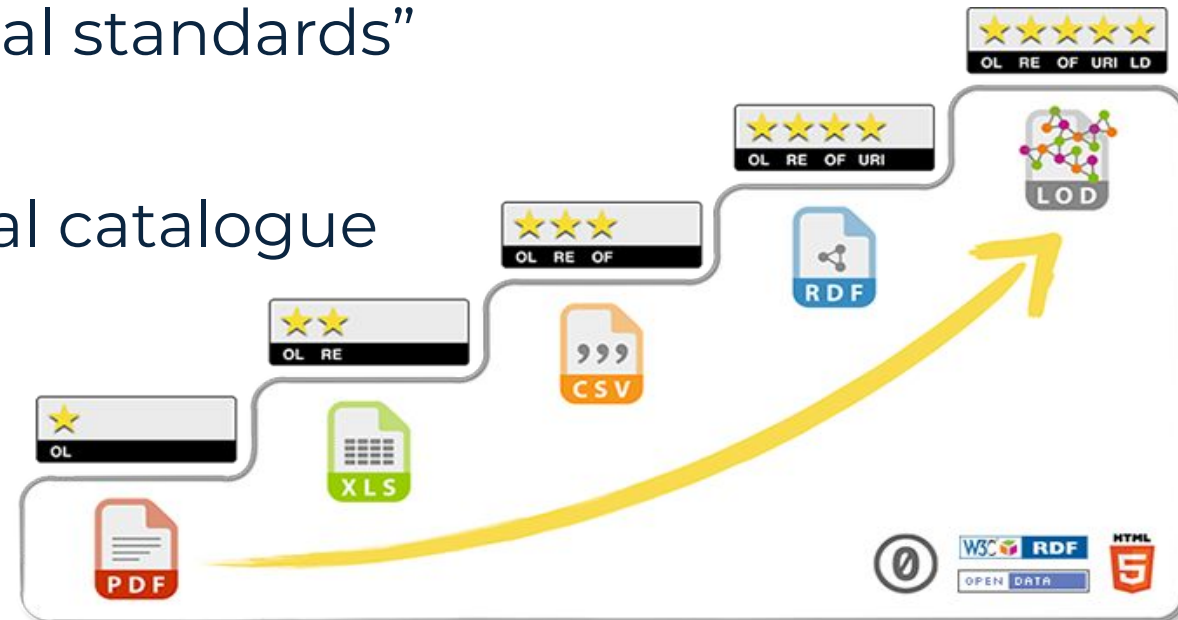
/// The big picture

- Local catalogue of open data - CKAN
 - <https://opendata.mzcr.cz/>
 - 65 datasets



/// The big picture

- datasets distributions conforms to CSVW (CSV on the Web)
 - dataset.csv
 - dataset.csv-metadata.json
- data types follow “Open formal standards”
 - <https://ofn.gov.cz/>
- datasets are published to local catalogue



/// Publishing process



Publishing process

Data harvesting

- Various data sources
 - database dumps
 - REST APIs
 - static files



Publishing process

Data processing

- Data specific operations if needed
 - Aggregation
 - Normalization
 - Anonymization
- Saving data to final CSV distribution
- Mapping dataset fields with JSON schema

/// Publishing process

REST API
DB DUMP
CSV

harvesting



```
[  
  {  
    "id" : 123,  
    "name" : "Apple",  
    "nick" : "Miloš"  
  },  
  {...},  
  {...}  
]
```

processing



csv + json schema



Publishing process

Validation

- Several open-source tools to validate locally
 - rdf-tabular (<https://github.com/ruby-rdf/rdf-tabular>)
 - csvlint (<https://github.com/Data-Liberation-Front/csvlint.rb>)
 - appears to be dead :-(
 - csvw-validator (<https://github.com/malyvoj3/csvw-validator>)
- ... or online
 - <https://csvw.opendata.cz/validation>



Publishing process

Dataset publishing

- Store distribution in accessible destination
- Update local OD catalogue



Technical solution && Use cases

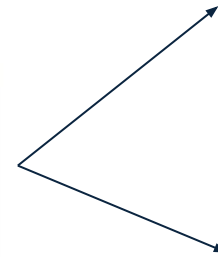


Technical solution

NRPZS dataset (National register of health services providers)

- Dataset must be published as OD due the legislation and updated monthly

**NRPZS
REST API**



csvw-validator



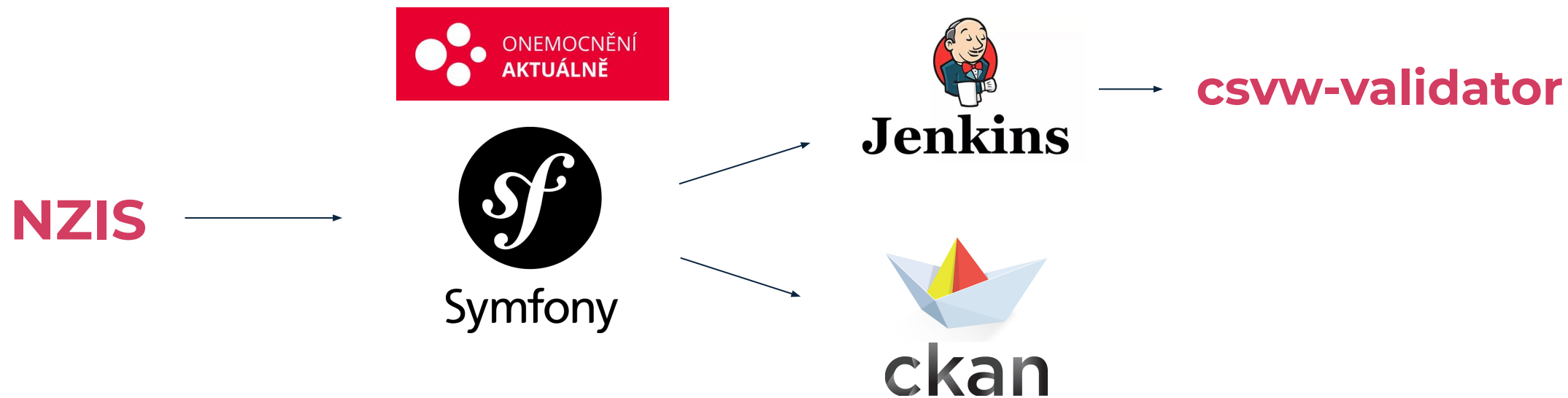
ckan





Technical solution

onemocneni-aktualne.cz/api/v2/covid-19



A group of five people are gathered in a modern office setting, likely for a meeting or collaborative work session. In the foreground, a woman with blonde hair, wearing a patterned cardigan, is gesturing with her hands as if explaining something. To her left, a man with glasses and a striped shirt is seated at a table with a laptop. Behind them, a woman with dark hair and a man with a beard are looking on. To the right, another woman with long dark hair is partially visible. The background shows office shelves and plants, creating a professional yet relaxed atmosphere. The entire image is overlaid with a semi-transparent blue filter.

/// Questions?

WEB /// STUDIO