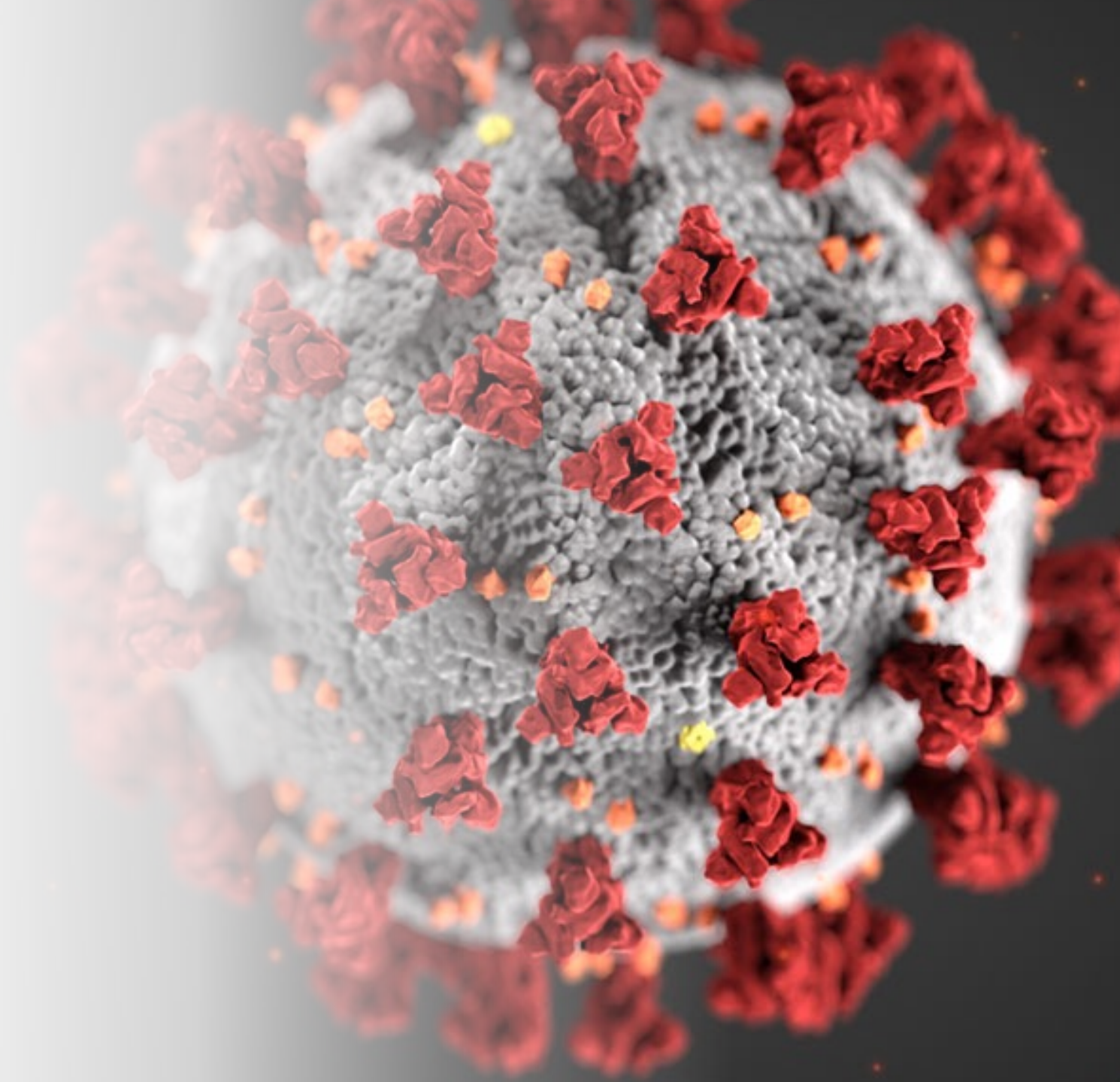


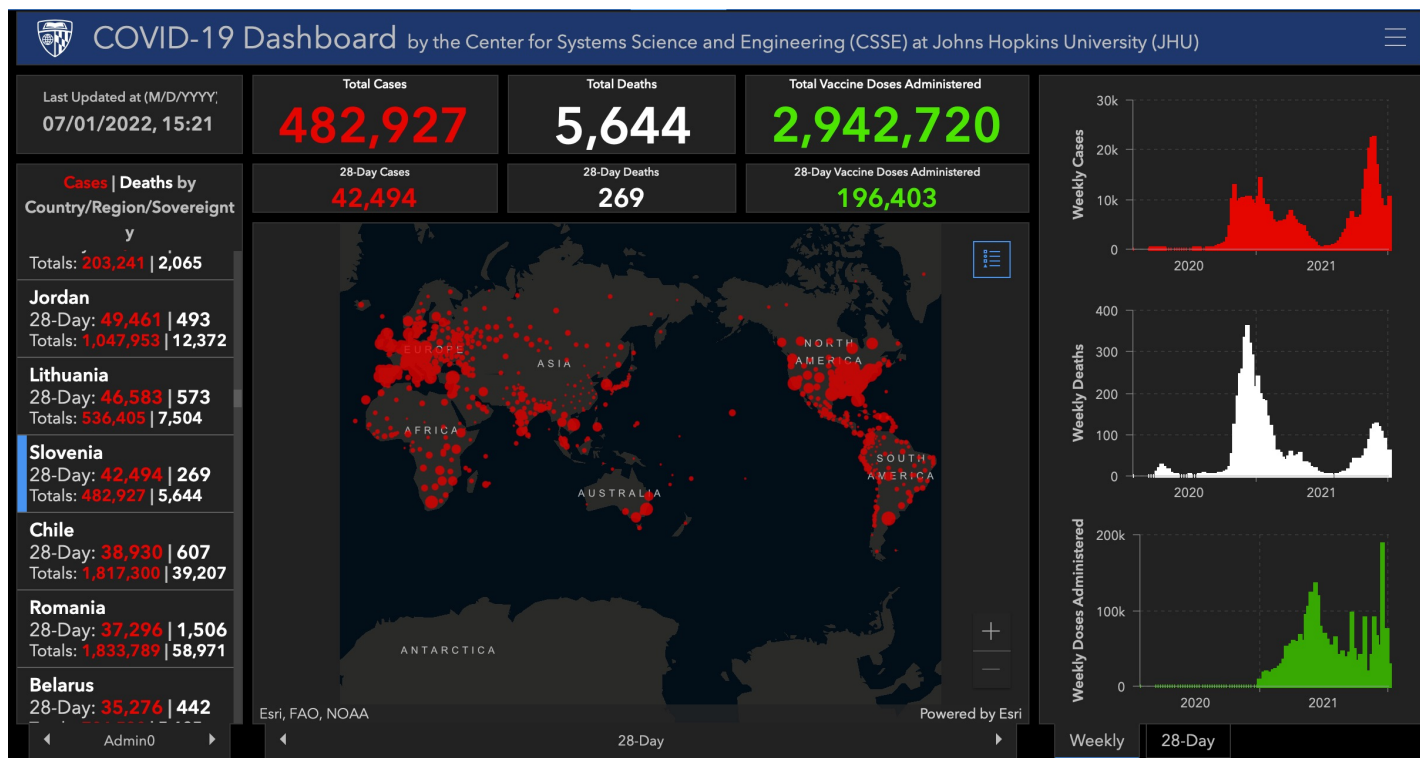
# Povezovanje COVID-19 podatkovnih zbirk

---

Eva Bizilj, Nikolay Vasilev



# Problemska domena



- podatkovne zbirke
  - Covid-19 sledilnik
  - JHU CSSE
  - REST Countries
- identifikacija atributov in podatkovni model
- podatkovna baza
- poizvedbe nad tabelami
- analiza in vizualizacija v RStudios

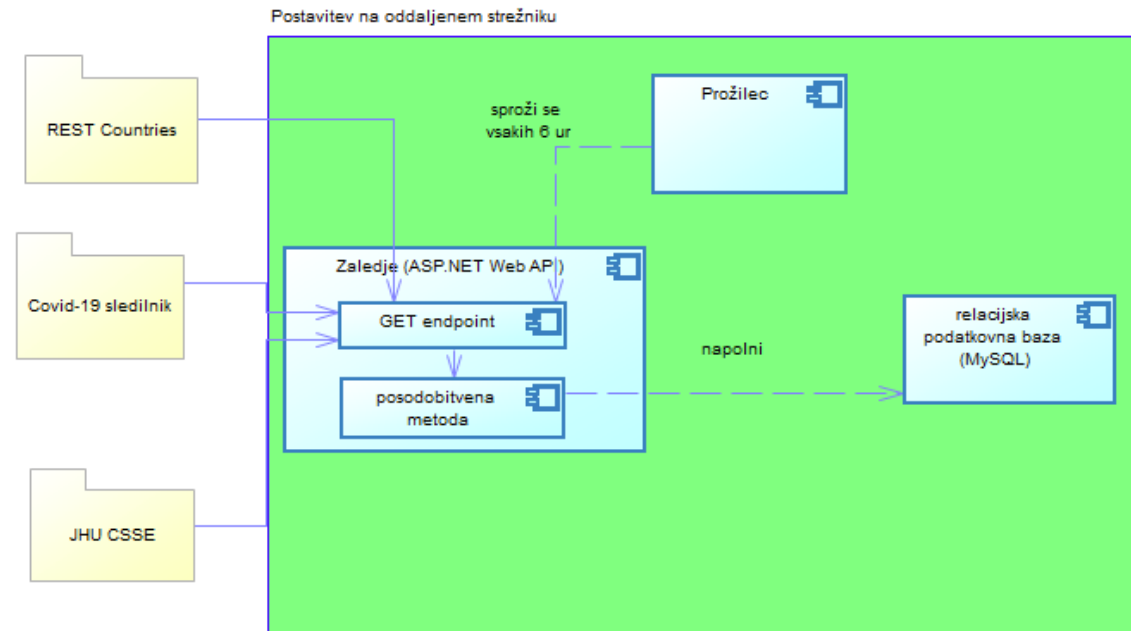
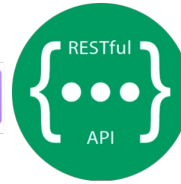
# Težave s podatki in podatkovnimi zbirkami

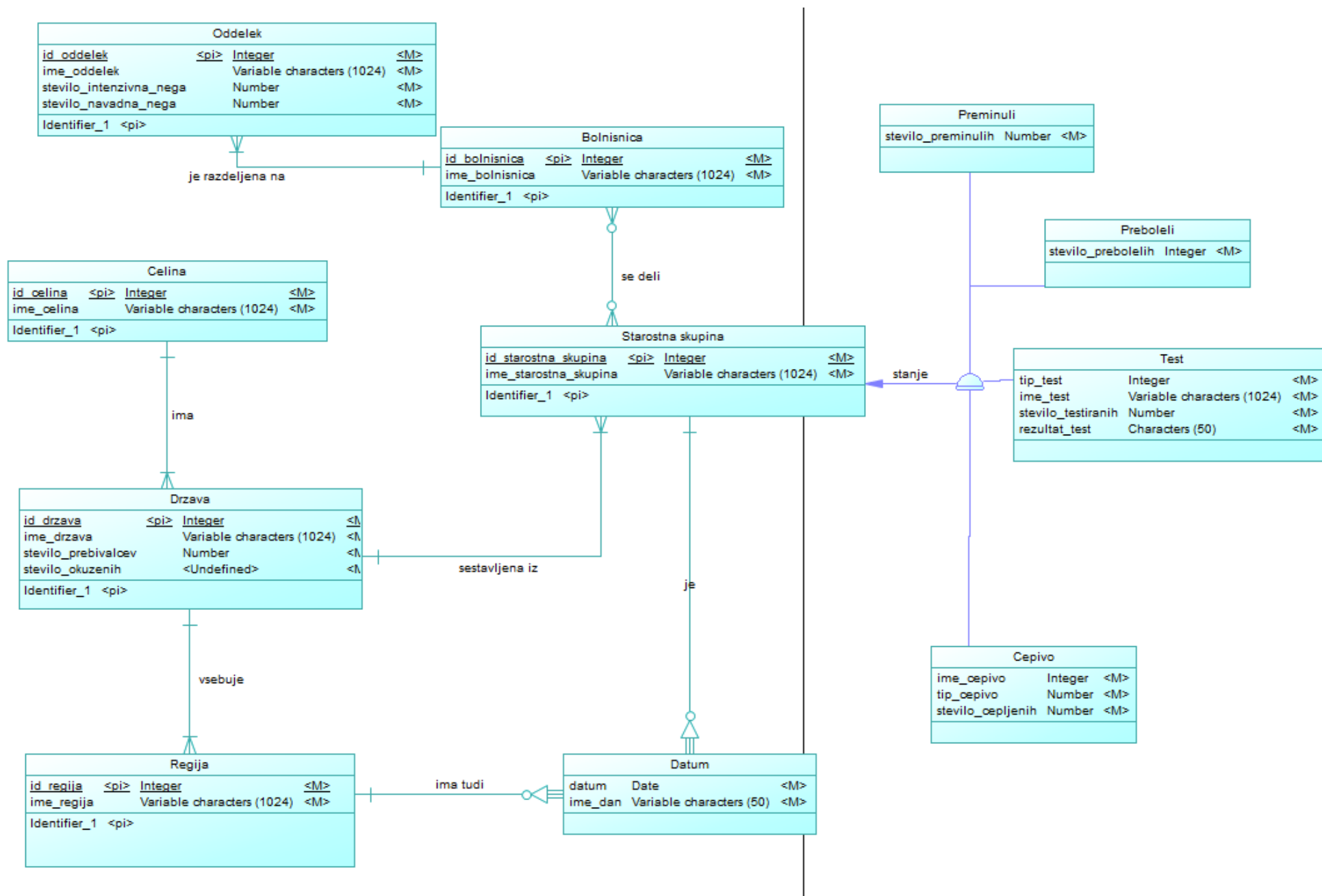
- neurejenost JSON podatkov (slaba JSON struktura)
- manjkajoči podatki
- premalo podatkov (samo za neko kratko časovno obdobje)
- neuporabni podatki (npr. samo število smrti na 100.000 prebivalcev in ne dejansko število smrti)

težave z razčlenjevanjem ("parsanje") podatkov in obravnavo napak na zaledju (*backend*)

```
{
  "records" : [
    {
      "dateRep" : "14/12/2020",
      "day" : "14",
      "month" : "12",
      "year" : "2020",
      "cases" : 746,
      "deaths" : 6,
      "countriesAndTerritories" : "Afghanistan",
      "geoId" : "AF",
      "countryterritoryCode" : "AFG",
      "popData2019" : 38041757,
      "continentExp" : "Asia",
      "Cumulative_number_for_14_days_of_COVID-19_cases_per_100000" : "9.01377925"
    },
    {
      "dateRep" : "13/12/2020",
      "day" : "13",
      "month" : "12",
      "year" : "2020",
      "cases" : 298,
      "deaths" : 9,
      "countriesAndTerritories" : "Afghanistan",
      "geoId" : "AF",
      "countryterritoryCode" : "AFG",
      "popData2019" : 38041757,
      "continentExp" : "Asia",
      "Cumulative_number_for_14_days_of_COVID-19_cases_per_100000" : "7.05277624"
    }
  ]
}
```

```
1 {
2   "date": "2021-01-10T00:00:00",
3   "region": [
4     {
5       "regionName": "lj",
6       "dailyActiveCases": 5282,
7       "deceasedToDate": 722,
8       "firstVaccineToDate": 8985,
9       "secondVaccineToDate": 1
10    }
11  ]
12 },
13 {
14   "date": "2021-01-11T00:00:00",
15   "region": [
16     {
17       "regionName": "lj",
18       "dailyActiveCases": 5146,
19       "deceasedToDate": 727,
20       "firstVaccineToDate": 9327,
21     }
22   ]
23 }
```







```
sledilnikCovid.Infrastructure | sledilnikCovid.Infrastructure.Implementation.SledilnikDataClient | client
4 using System.Collections.Generic;
5 using System.Linq;
6 using System.Text;
7 using System.Threading.Tasks;
8
9 namespace sledilnikCovid.Infrastructure.Implementation
10 {
11     1 reference
12     public class SledilnikDataClient : IFormatFetcher
13     {
14         private static readonly HttpClient client = new HttpClient();
15
16         1 reference
17         public async Task<List<string>> getCSV()
18         {
19             var response = await client.GetAsync("https://raw.githubusercontent.com/sledilnik/data/master/csv/region-cases.csv");
20             if (response.IsSuccessStatusCode)
21             {
22                 string responseBody = await response.Content.ReadAsStringAsync();
23                 List<string> responseList = responseBody.Split('\n').ToList();
24                 responseList = responseList.Take(responseList.Count() - 1).ToList();
25
26                 return responseList;
27             }
28             else
29             {
30                 return null;
31             }
32         }
33     }
34 }
```

```
/// <summary>
/// Vrne vse primere z informacijami o podatkih osnovanih na datumu in regiji
/// </summary>
/// <param name="region"></param>
/// <param name="from"></param>
/// <param name="to"></param>
/// <returns>Vrne seznam primerov</returns>
/// <response code="200">Vrne seznam primerov glede na željeno regijo v dani državi</response>
/// <response code="400">Vrne izjemo glede na neustrezne vhodne podatke uporabnika</response>
[HttpGet]
[Route("slovenia/cases")]
[ProducesResponseType(StatusCodes.Status200OK)]
[ProducesResponseType(StatusCodes.Status400BadRequest)]
0 references
public async Task<ActionResult<List<CasesDto>>> GetCases(string? country, string? region, DateTime
{
    //collect syntax and semantic errors
    var validRegions = new List<string> { "lj", "ce", "kr", "nm", "kk", "kp", "mb", "ms", "ng", "po" };
    var exceptionStack = new List<string>();
    DateTime dDate;

    if (region != null && !validRegions.Contains(region))
        exceptionStack.Add("Invalid region.");

    if (from > to)
        exceptionStack.Add("'From' date cannot be later than 'To' date*");

    //evaluate syntax and semantic errors
    if (exceptionStack.Count == 0)
    {
        var data = await _regionService.FetchDataCases(region, from, to);
        return Ok(data);
    }
    else {
        return BadRequest(exceptionStack);
    }
}
```

https://localhost:7238/api/slovenia/cases?region=lj&from=1%2F10%2F2021&to=3%2F10%2F2021

GET https://localhost:7238/api/slovenia/cases?region=lj&from=1%2F10%2F2021&to=3%2F10%2F2021

Params Authorization Headers (8) Body Pre-request Script Tests Settings

Body Cookies Headers (4) Test Results Status: 200 OK Time: 188 ms Size: 9.58 KB

Pretty Raw Preview Visualize JSON

```
1 [
2   {
3     "date": "2021-01-10T00:00:00",
4     "region": [
5       {
6         "regionName": "lj",
7         "dailyActiveCases": 5282,
8         "deceasedToDate": 722,
9         "firstVaccineToDate": 8985,
10        "secondVaccineToDate": 1
11      }
12    ]
13  },
14  {
15    "date": "2021-01-11T00:00:00",
16    "region": [
17      {
18        "regionName": "lj",
19        "dailyActiveCases": 5146,
20        "deceasedToDate": 727,
21        "firstVaccineToDate": 9327,
```

17. Vrne število okuženih, število hospitalizacij na navadni in intenzivni negi ter število smrti v obdobju, ko je prevladovala originalna različica virusa Covid-19 v Sloveniji.

```
SELECT t5.STEVILO_POZITIVNIH, t5.STEVILO_INTENZIVNA_NEGA, t5.STEVILO_NAVADNA_NEGA,
SUM(p.STEVILO_PREMINULIH) AS STEVILO_SMRTI
FROM PREMINULI p
    JOIN (
        SELECT t4.ID_STAROSTNA_SKUPINA, t4.STEVILO_INTENZIVNA_NEGA, t4.STEVILO_NAVADNA_NEGA,
SUM(ts.STEVILO_TESTIRANIH) AS STEVILO_POZITIVNIH
        FROM TEST ts
    JOIN (
        SELECT t3.ID_STAROSTNA_SKUPINA, o.STEVILO_INTENZIVNA_NEGA, o.STEVILO_NAVADNA_NEGA
        FROM ODDELEK o
    JOIN (
        SELECT t2.ID_STAROSTNA_SKUPINA, b.ID_BOLNISNICA
        FROM BOLNISNICA b
    JOIN (
        SELECT t1.ID_STAROSTNA_SKUPINA, sd.ID_BOLNISNICA
        FROM SE_DELI sd
    JOIN (
        SELECT ss.ID_STAROSTNA_SKUPINA
        FROM DATUM d
    JOIN STAROSTNA_SKUPINA ss
        ON d.ID_STAROSTNA_SKUPINA = ss.ID_STAROSTNA_SKUPINA
        WHERE d.DATUM >= '2020-03-04'
        AND d.DATUM <= '2021-04-01'
    ) t1
    ON sd.ID_STAROSTNA_SKUPINA = t1.ID_STAROSTNA_SKUPINA
    ) t2
    ON t2.ID_BOLNISNICA = b.ID_BOLNISNICA
    ) t3
    ON o.ID_BOLNISNICA = t3.ID_BOLNISNICA
    ) t4
```

```
ON ts.ID_STAROSTNA_SKUPINA = t4.ID_STAROSTNA_SKUPINA
WHERE ts.REZULTAT_TEST="positive"
) t5
ON p.ID_STAROSTNA_SKUPINA = t5.ID_STAROSTNA_SKUPINA;
```

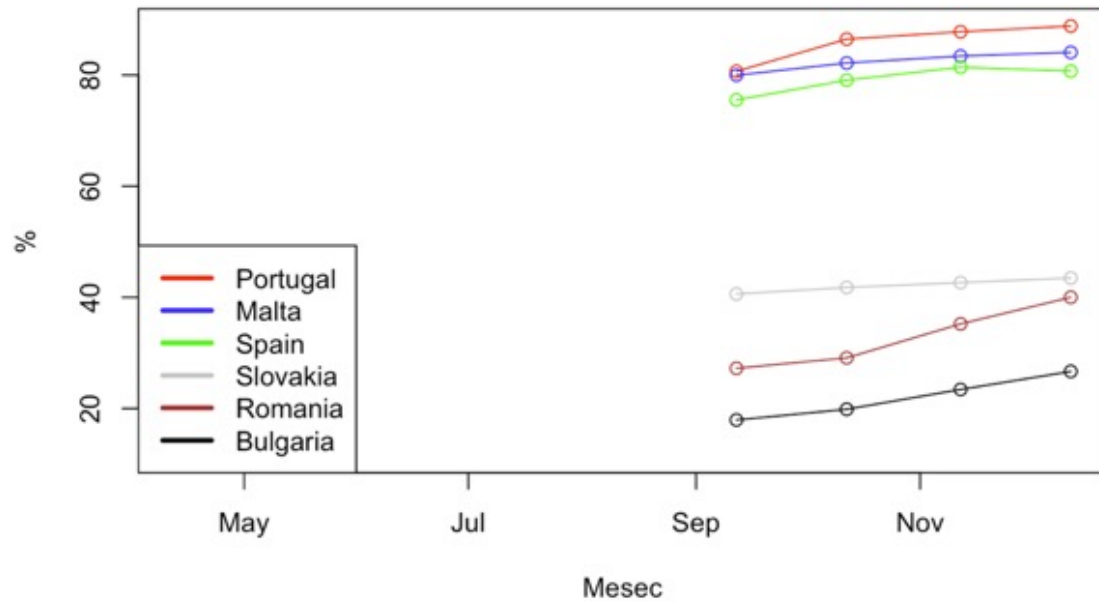
## Rezultat

poizvedbaC17.csv ×

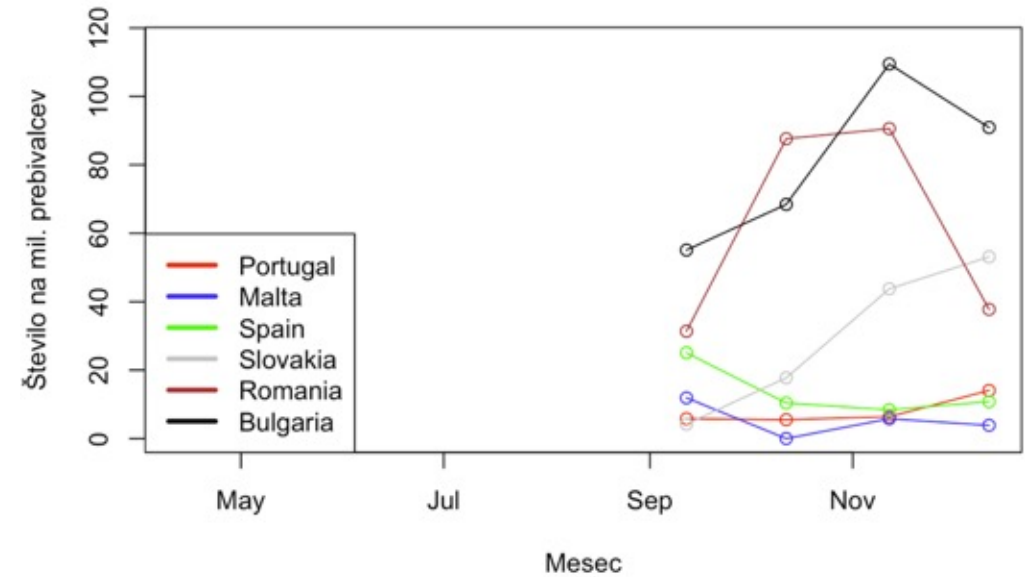
1	STEVILO_POZITIVNIH, STEVILO_INTENZIVNA_NEGA, STEVILO_NAVADNA_NEGA, STEVILO_SMRTI
2	218266, 27498, 136709, 4054



Delež cepljenih (zadnje 3 mesece leta 2021)



Število bolnikov na intenzivni negi (zadnje 3 mesece leta 2021)



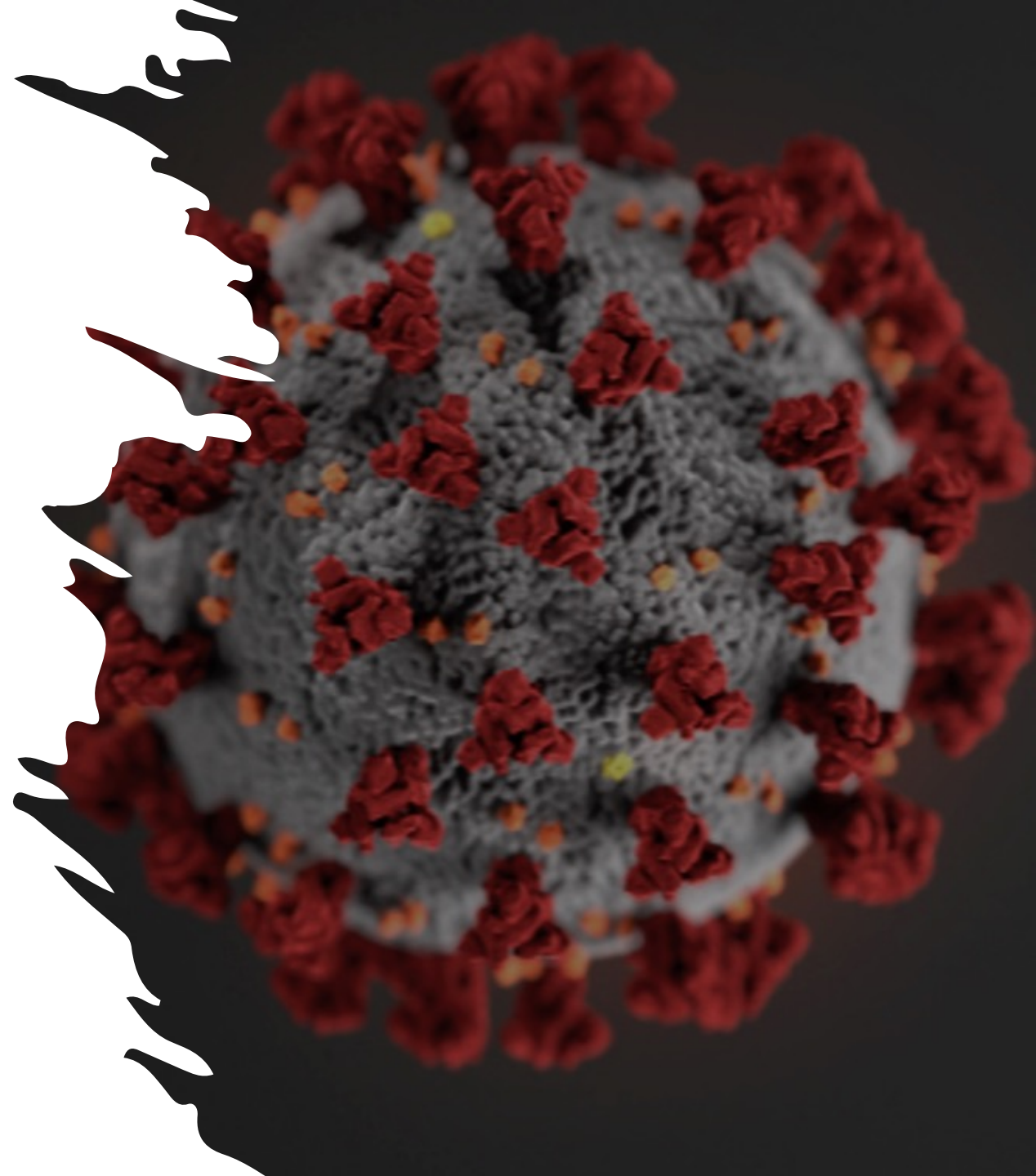
## Predpostavka 3

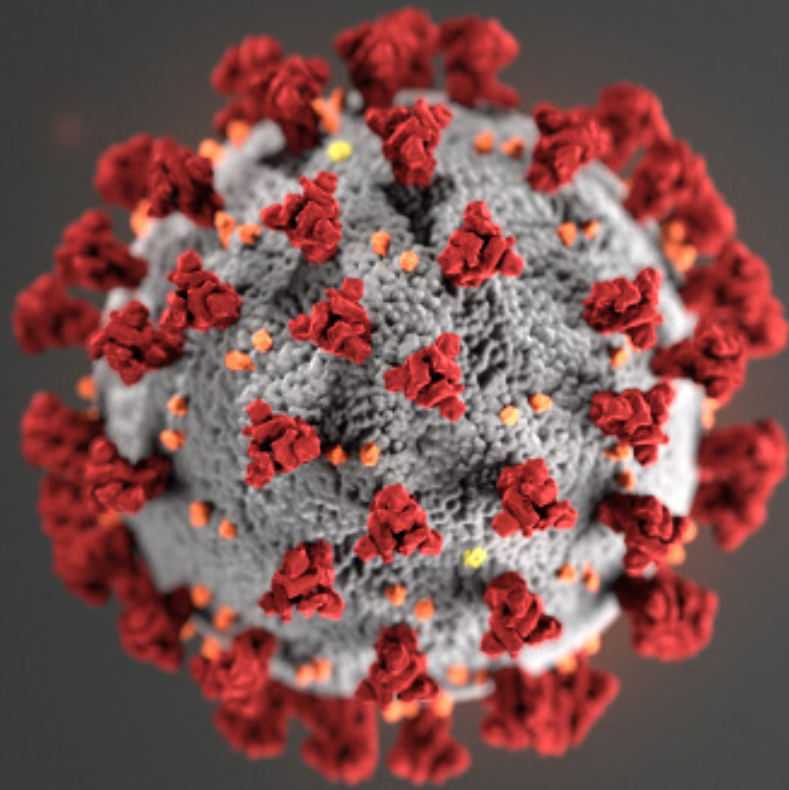
V državah, kjer je večji delež cepljenih, je manjši delež hospitalizacij na intenzivni negi.



# Viri

- <https://covid-19.sledilnik.org/sl/stats>
- <https://coronavirus.jhu.edu>
- <https://restcountries.com>





Hvala za pozornost!