UNIVERSITY OF CAMBRIDGE

PBSP

European Research Council
Established by the European Commission

HEP STATISTICS IN LATIN AMERICA

MANUEL MORALES ALVARADO

UNIVERSITY OF CAMBRIDGE

PBSP

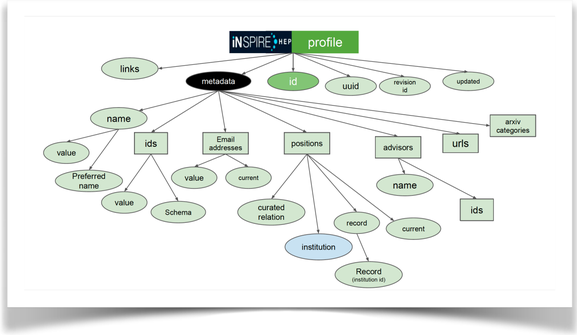
European Research Council
Established by the European Commission

HEP STATISTICS IN LATIN AMERICA

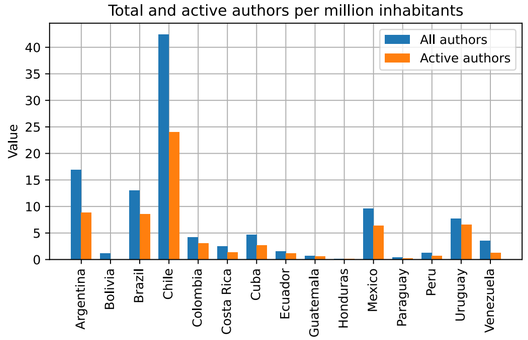
MANUEL MORALES ALVARADO

Thanks to D. Restrepo (U. de Antioquia) for some of the slides!

OUTLINE



Organisation, INSPIRE-HEP API, analysis code



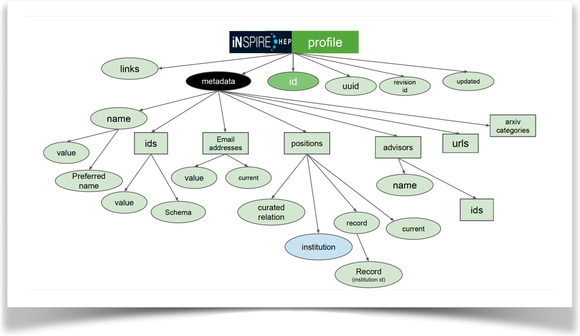
Results

Conclusions and outlook

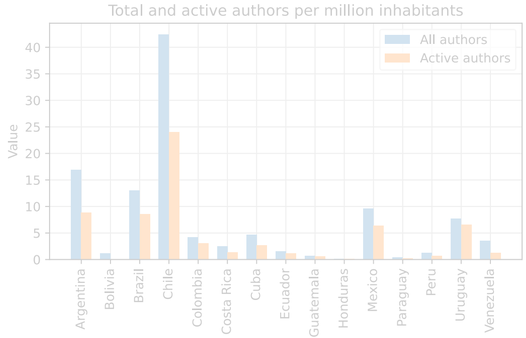
OBJECTIVE

Generate a public database, analysis, and report to assess the status of HEP research and researchers in Latin America

OUTLINE



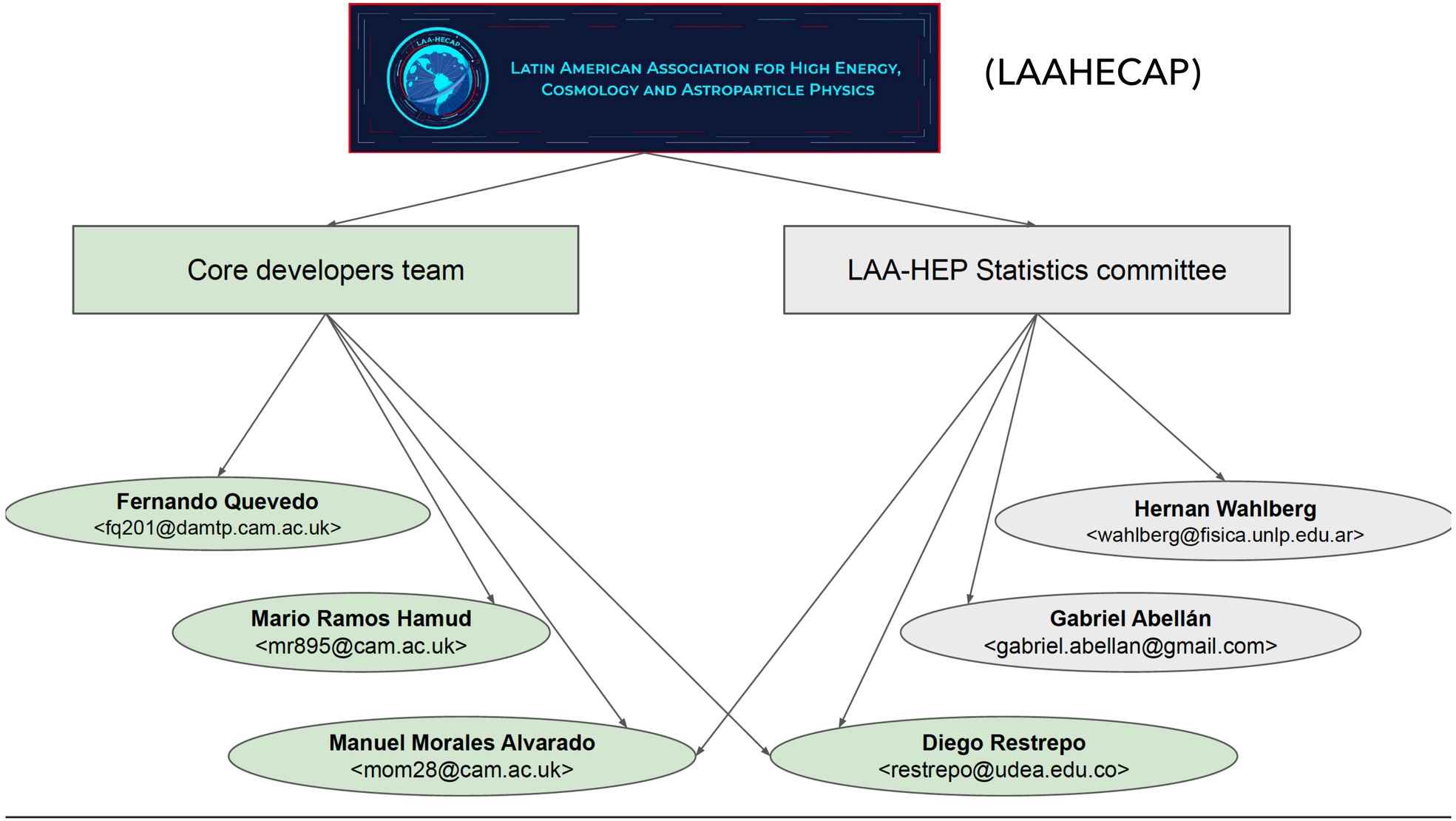
Organisation, INSPIRE-HEP API, analysis code



Results

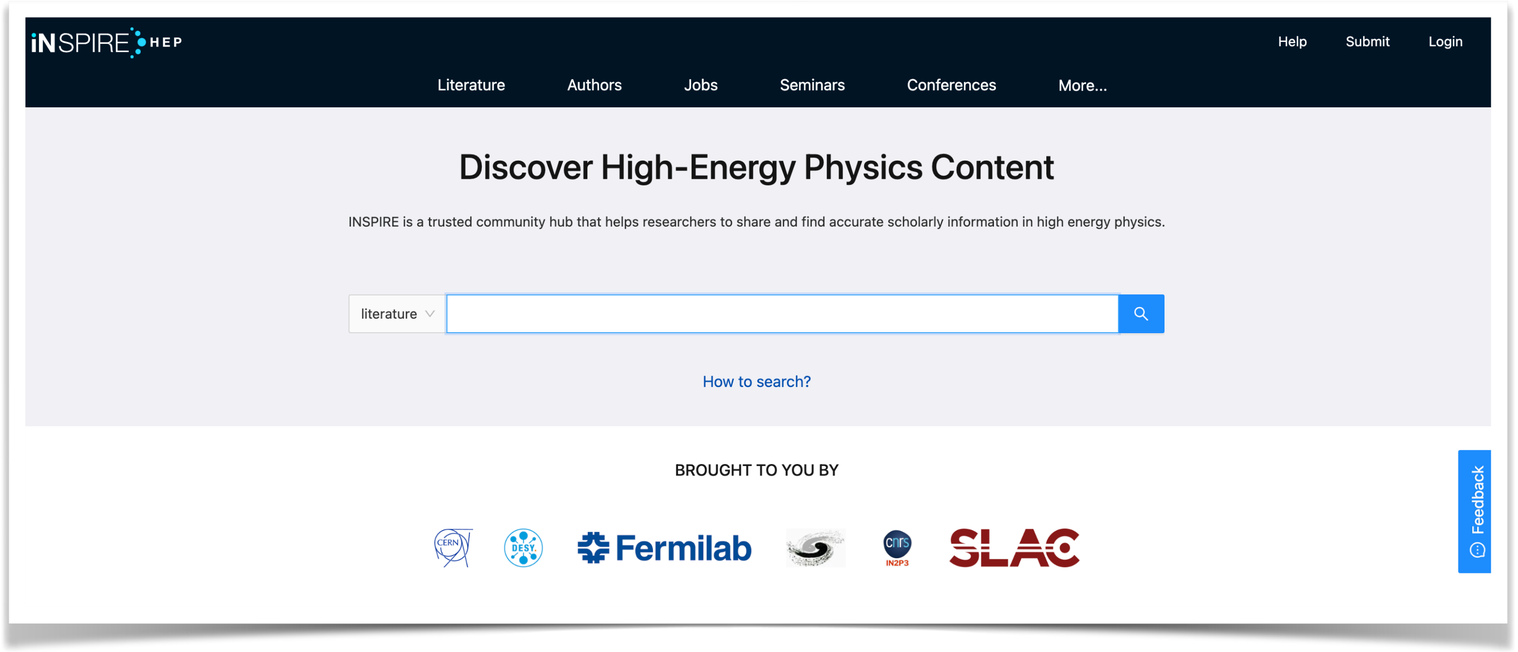
Conclusions and outlook

ORGANISATION



INSPIRE-HEP

- We use the https://inspirehep.net/ data for this analysis
- INSPIRE-HEP is an open access database
- First design conceived in 1967: SPIRES (Stanford Physics Information Retrieval System)
- SPIRES-HEP: the first website in America, and the first WWW dabatase



DATA STRUCTURES

Lists = []

```
[2]: l=['cat','dog','cow']
     l[1]
```

```
[2]: 'dog'
```

Dictionaries = {key, value}

```
[3]: d={'animal':'cat','fruit':'mango'}
     d['fruit']
```

```
[3]: 'mango'
```

They can be nested!

```
[4]: ds={'animals':l,'fruits':['mango','banana']}
     ds['animals'][1]
```

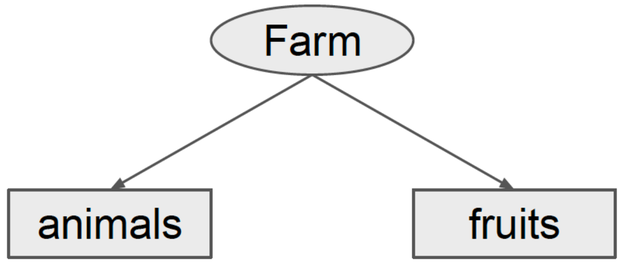
```
[4]: 'dog'
```

```
[5]: from IPython import display
```

```
[6]: display.JSON(ds)
```

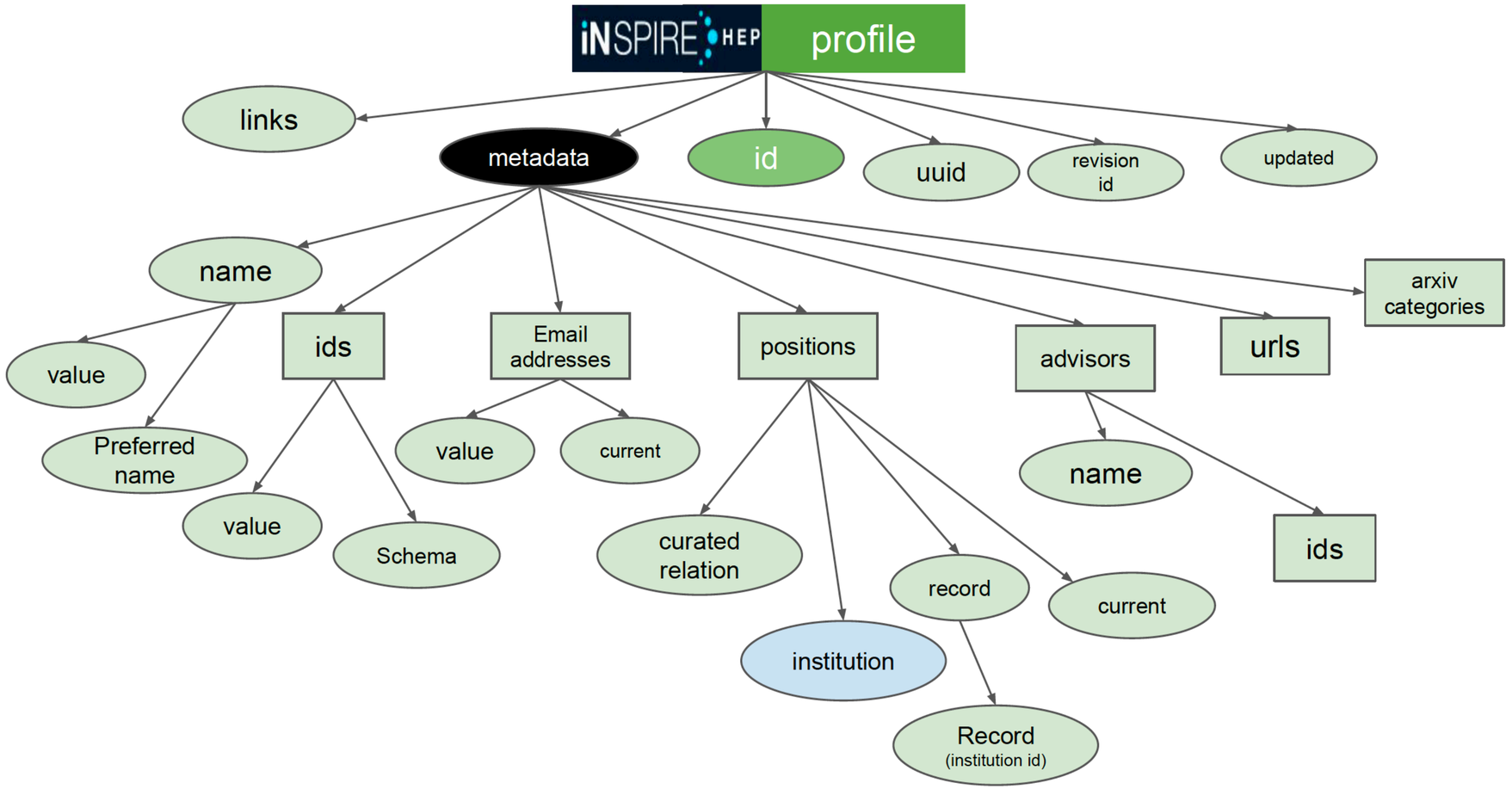
```
[6]: ▾ root:
       ▾ animals: []  3 items
           0: "cat"
           1: "dog"
           2: "cow"
       ▸ fruits: []  2 items
```

```
[ ]:
```



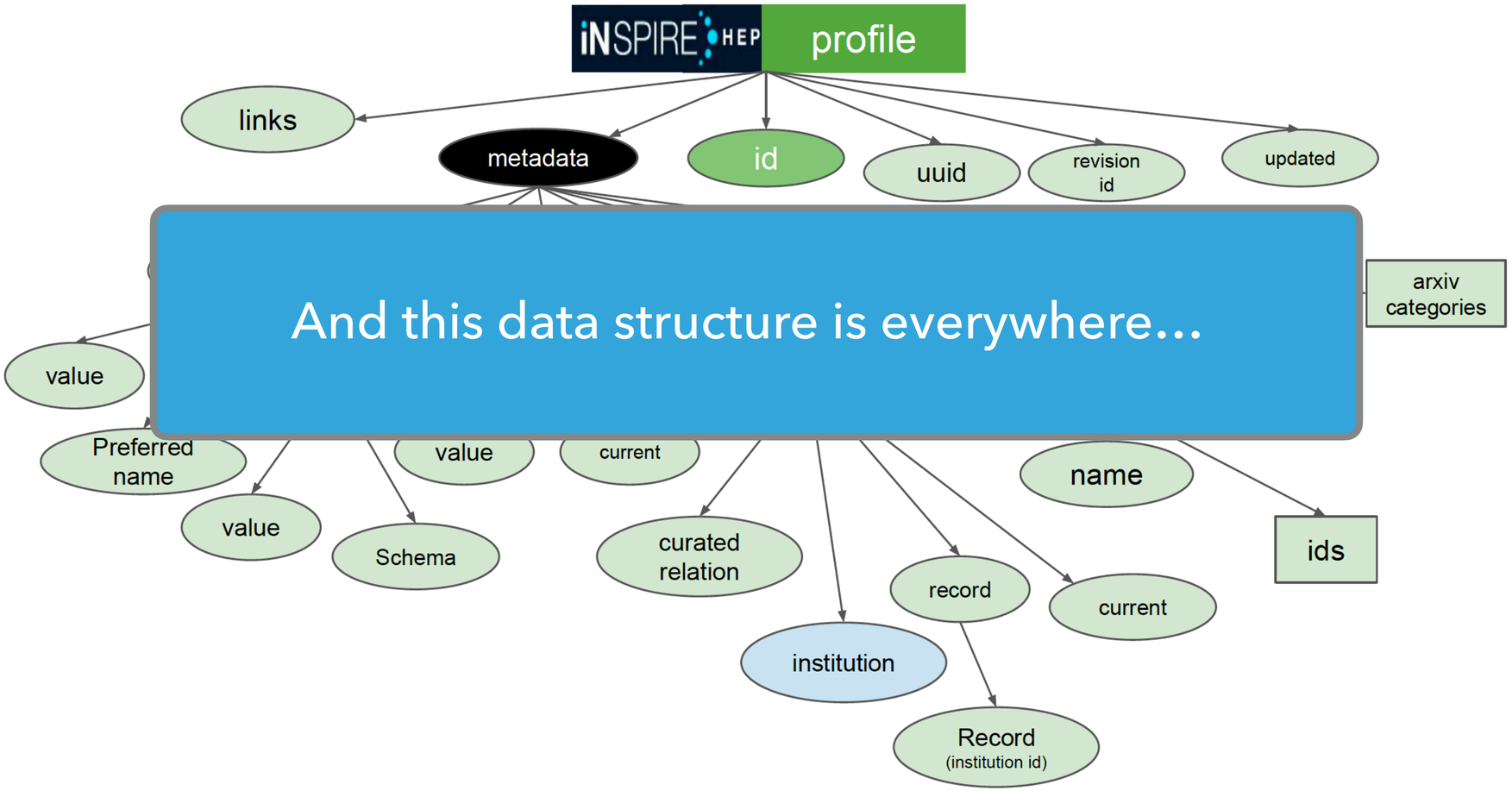
INSPIRE-HEP APPLICATION PROGRAMMING INTERFACE (API)

A profile on INSPIRE-HEP gives us access to different information



INSPIRE-HEP API

A profile on INSPIRE-HEP gives us access to different information

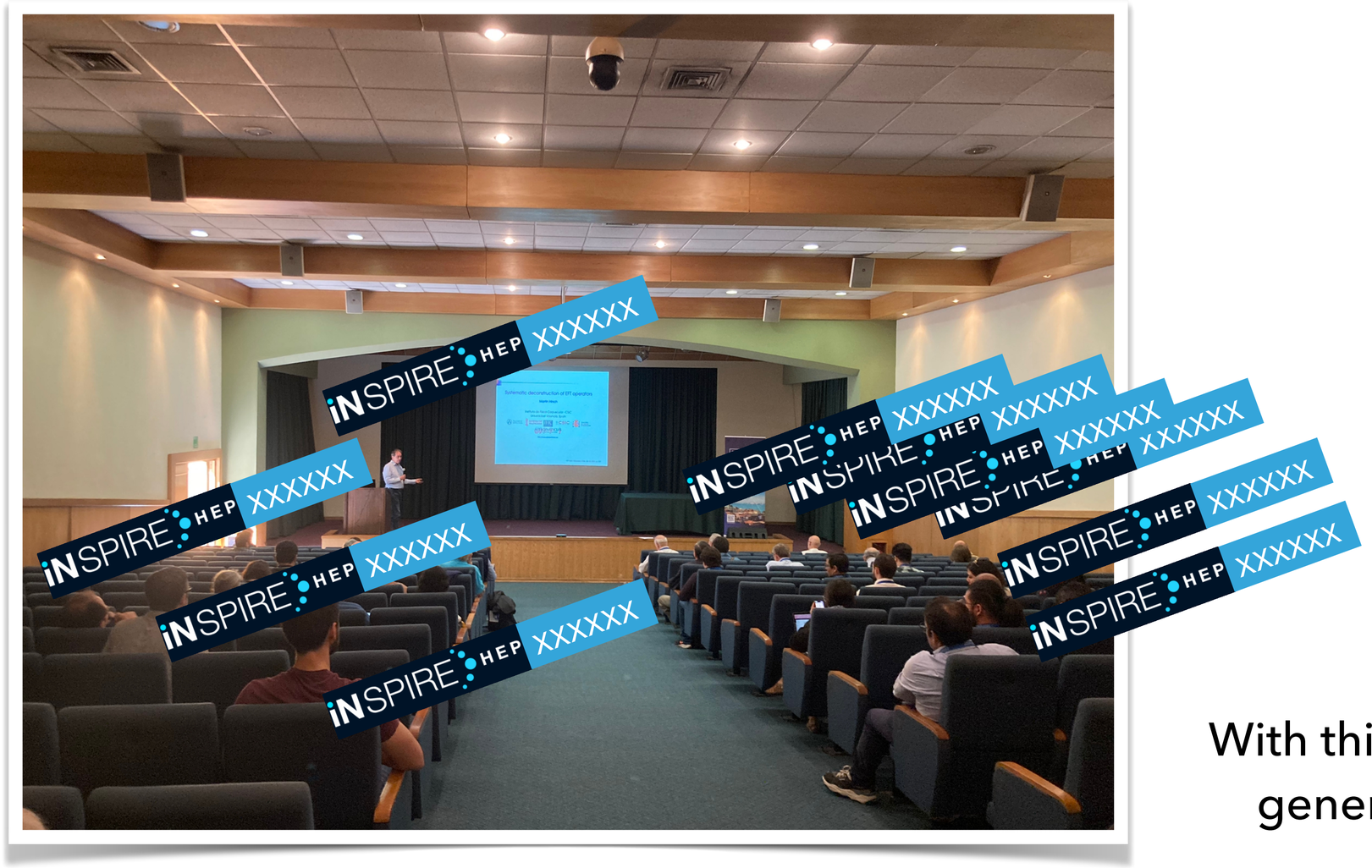


INSPIRE-HEP API FOR THE PHYSICS COMMUNITY

This is us

INSPIRE-HEP API FOR THE PHYSICS COMMUNITY

This is us

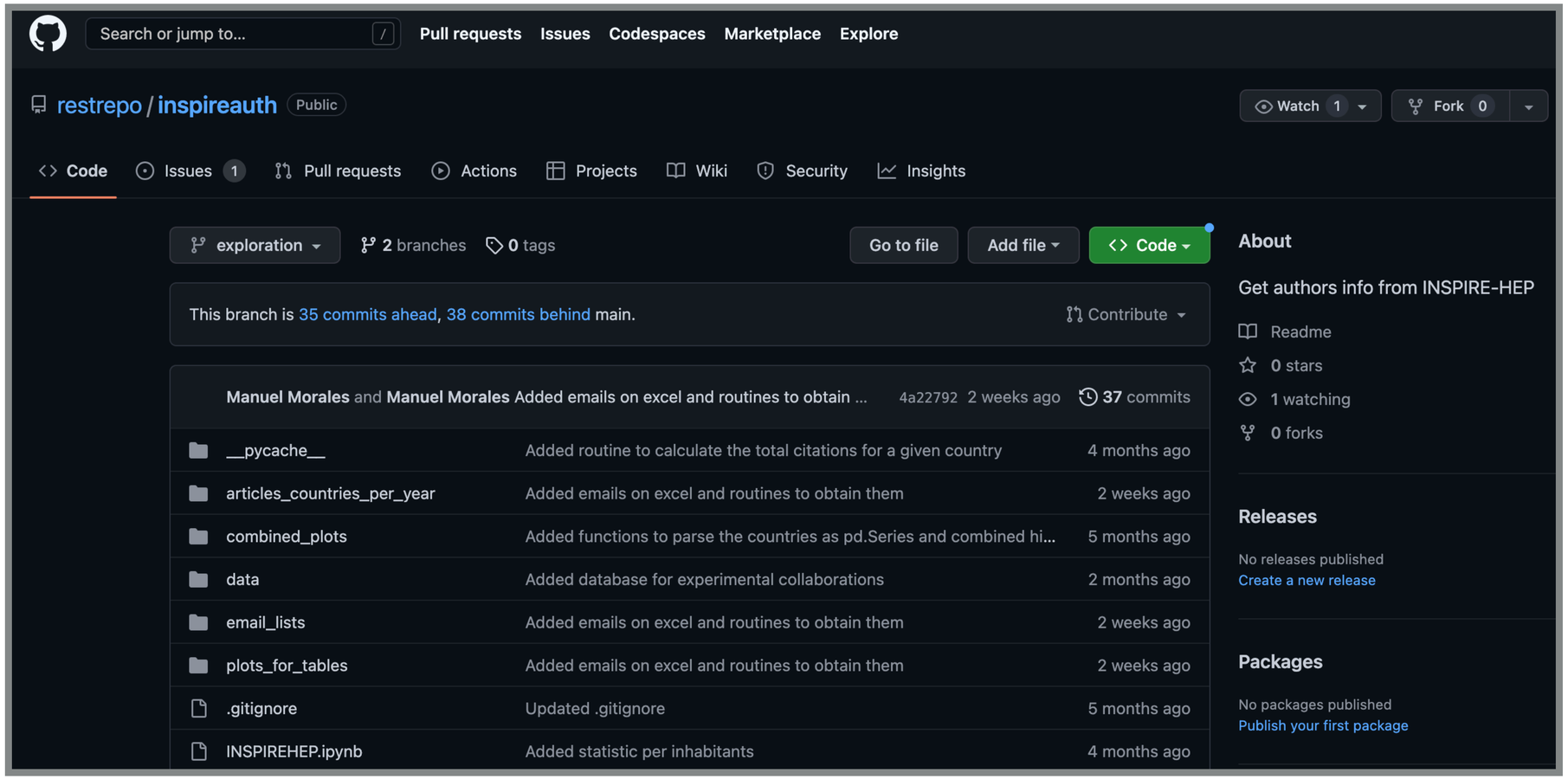


With this we have generated ...

DATABASE AND CODE

Completely open source!

https://github.com/restrepo/inspireauth/tree/main

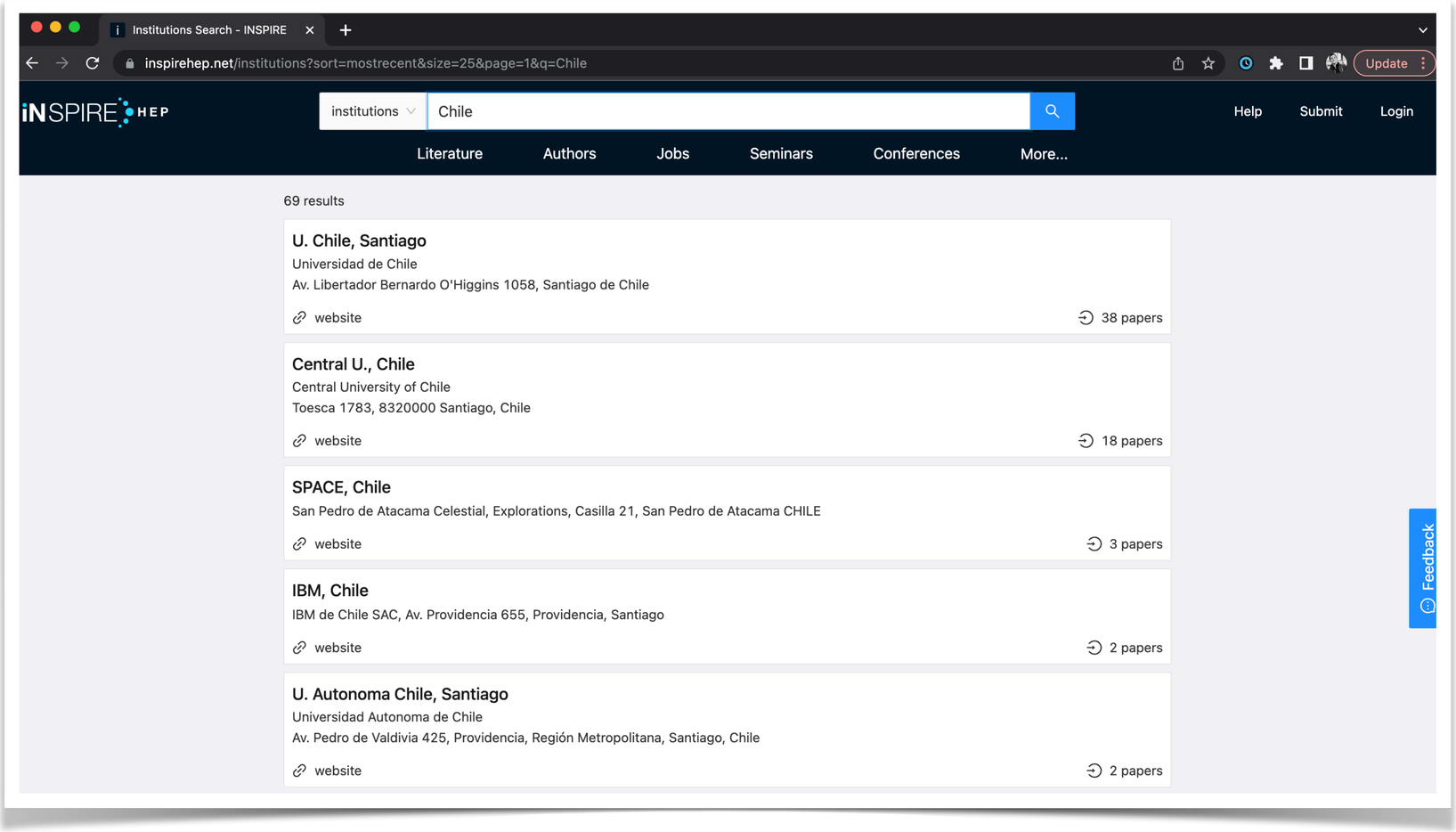


A short overview of the logic of the database...

A SHORT EXERCISE

“Extract the list of all the authors from Chile from INSPIRE-HEP”

1 - Access the institutions from Chile

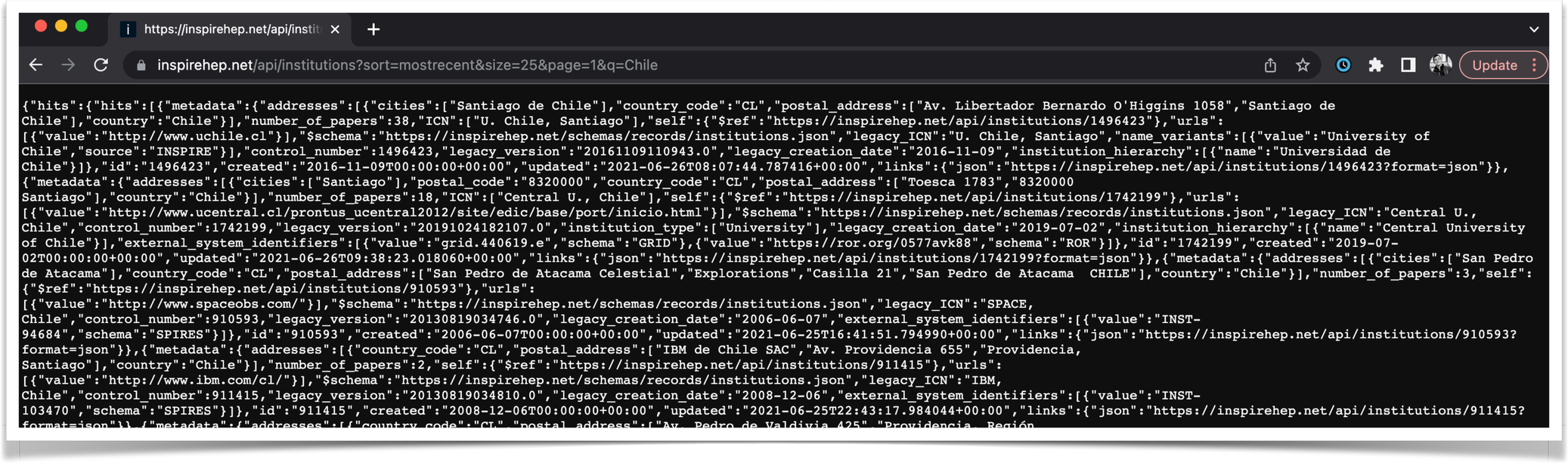


A SHORT EXERCISE

2 - Access the API

https://inspirehep.net/institutions?sort=mostrecent&size=25&page=1&q=Chile

https://inspirehep.net/api/institutions?sort=mostrecent&size=25&page=1&q=Chile



A SHORT EXERCISE

3 - Extract the list of institutions of Chile

```
In [35]: 1 list_legacies_ICN

Out[35]: ['U. Chile, Santiago',
          'Central U., Chile',
          'IBM, Chile',
          'SPACE, Chile',
          'European Southern Obs., Chile',
          'U. Autonoma Chile, Santiago',
          'San Sebastian U., Chile',
          'Santiago de Chile U.',
          'Unlisted, CL',
          'Pontificia U. Catol. Chile, Santiago',
          'Chile U., CMM',
          'Chile U., Beauchef',
          'Chile U., Santiago',
          'Chile Austral U., Valdivia',
          'Chile U., Catolica',
          'U. Arturo Prat, Iquique',
          'INI, Santiago',
          'CIDCA, Santiago',
          'Valparaiso U.',
          'Cerro Calan Observ.',
          'Concepcion Catolica U.',
          'Catolica del Norte U.',
          'Valparaiso U., Catolica',
          'CECS, Valdivia',
          'Temuco U., Catolica',
```

A SHORT EXERCISE

4 - Access the papers published by an institution. Then, iterate over institutions

https://inspirehep.net/api/literature?sort=mostrecent&page=1&q=aff+U.+Chile,+Santiago+and+ac+1->+10

aff: use `legacy_ICN`

and: logical operator

ac: defines number of authors

A SHORT EXERCISE

5 - For each article of each institution, extract the profile URL of the authors

In each entry:

File Edit View Insert Cell Kernel Help

Run Code

```
Out[152]: [['https://inspirehep.net/api/authors/1973718',
            'https://inspirehep.net/api/authors/1403047',
            'https://inspirehep.net/api/authors/1477164',
            'https://inspirehep.net/api/authors/2013953',
            'https://inspirehep.net/api/authors/1413660',
            'https://inspirehep.net/api/authors/999090'],
           ['https://inspirehep.net/api/authors/1993758',
            'https://inspirehep.net/api/authors/1993759',
            'https://inspirehep.net/api/authors/995374',
            'https://inspirehep.net/api/authors/1037240'],
           ['https://inspirehep.net/api/authors/1017586',
            'https://inspirehep.net/api/authors/1875153',
            'https://inspirehep.net/api/authors/996257',
            'https://inspirehep.net/api/authors/993295',
            'https://inspirehep.net/api/authors/1287459',
            'https://inspirehep.net/api/authors/985110'],
           ['https://inspirehep.net/api/authors/1037240',
            'https://inspirehep.net/api/authors/1987711',
            'https://inspirehep.net/api/authors/995374'],
```

We have obtained all the authors from Chile

SOME COMMENTS

- In this way we have obtained the list of all papers of a given country with up to 10 authors
- This means that the papers of experimental collaborations are left out of this analysis (we have another database for them!)
- The database was obtained in June 2022, and has to be updated after some time
- The actual nationality of the author is irrelevant. It counts as an author of a given country by being affiliated to an institution of that country

Now some results …

COUNTRIES IN THE DATABASE

We include the following countries in our study:

Country	Population
Argentina	45,606,000
Bolivia	11,833,000
Brazil	213,993,000
Chile	19,212,000
Colombia	51,266,000
Costa Rica	5,139,000
Cuba	11,318,000
Ecuador	17,888,000
Guatemala	18,250,000
Honduras	10,063,000
Mexico	130,262,000
Paraguay	7,220,000
Peru	33,359,000
Uruguay	3,485,000
Venezuela	28,705,000

The population of the countries is useful to find “metrics per capita”.

They have been obtained from https://data.un.org/en/index.html (checked in September 2022)

Why are some countries missing?

They simply do not have indexed entries on INSPIRE-HEP

METRICS FOR EACH COUNTRY

After generating the database we have obtained:

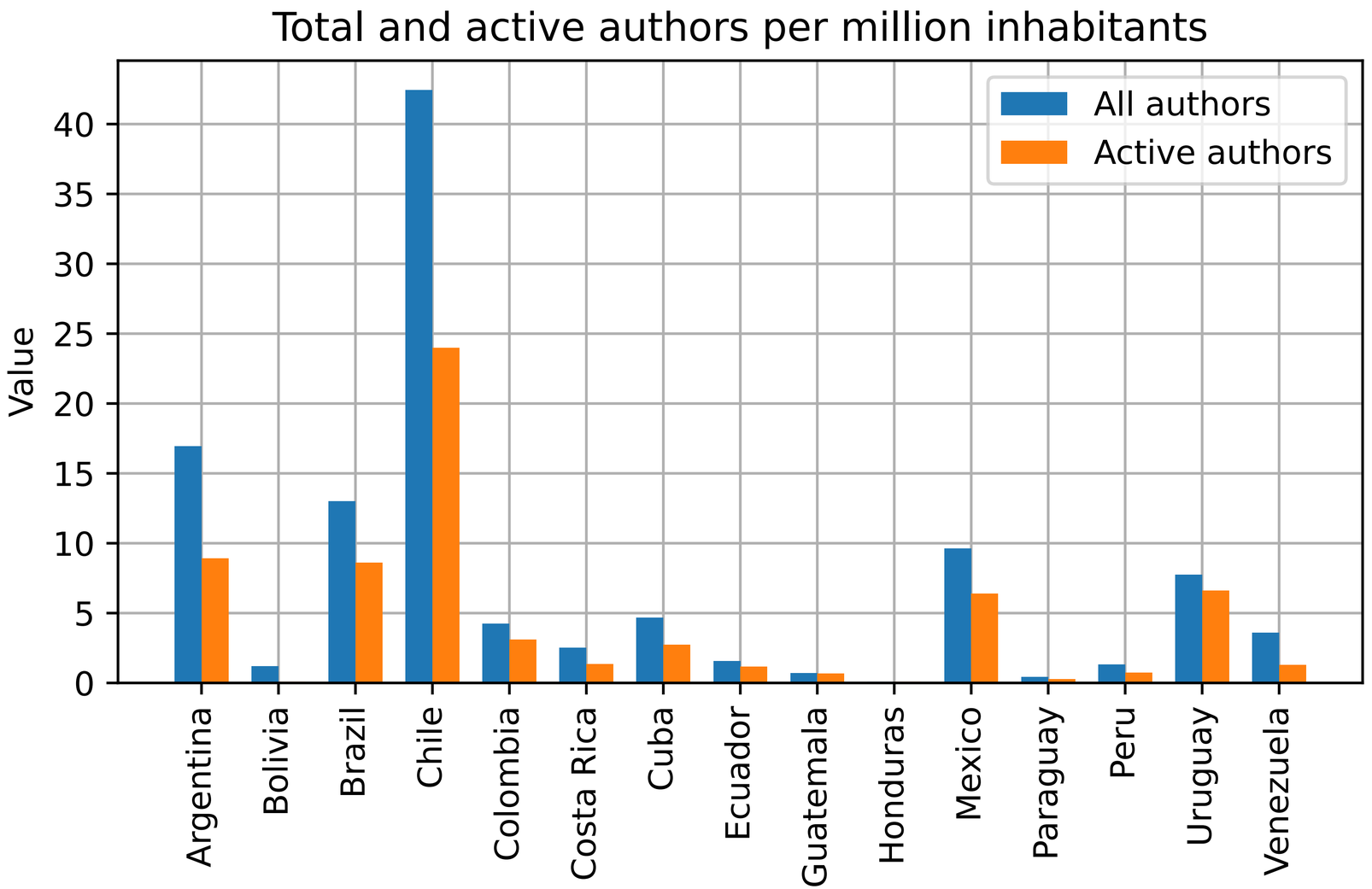
- Name and email of all active authors (with publications in the last 5 years)
- Number of total and active authors
- Number of publications
- Number of citations
- H-index
- Collaborations with other countries in Latin America
- And more...

TOTAL AND ACTIVE AUTHORS

(p.m.i. = per million inhabitants)

Country	Total authors	Active authors	Total authors p.m.i.	Active authors p.m.i
Argentina	773	406	17	9
Bolivia	14	1	1	0
Brazil	2782	1843	13	9
Chile	815	461	42	24
Colombia	218	159	4	3
Costa Rica	13	7	3	1
Cuba	53	31	5	3
Ecuador	28	21	2	1
Guatemala	13	12	1	1
Honduras	1	1	0	0
Mexico	1253	832	10	6
Paraguay	3	2	0	0
Peru	44	24	1	1
Uruguay	27	23	8	6
Venezuela	103	37	1	1

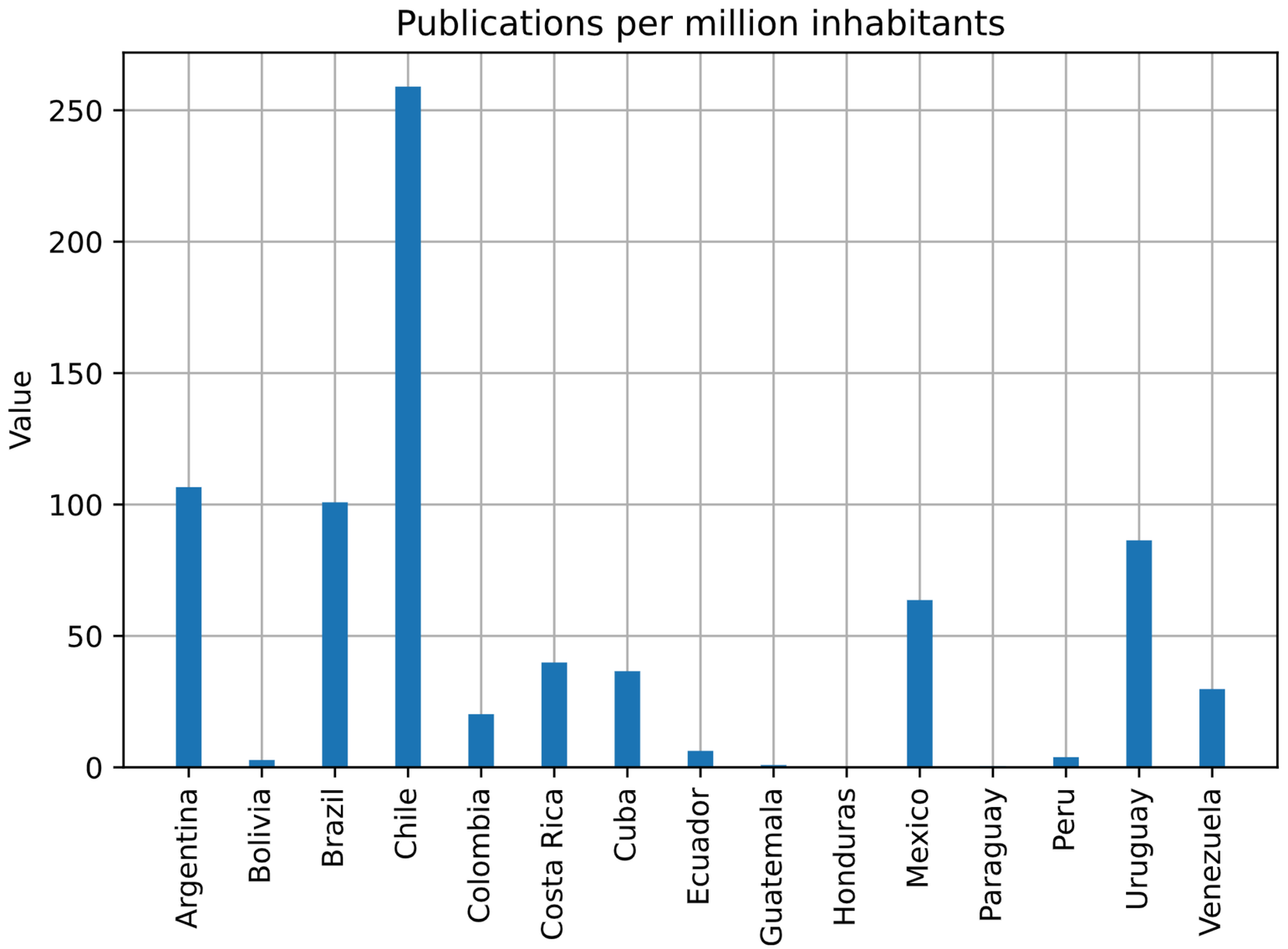
AUTHORS NORMALISED BY POPULATION



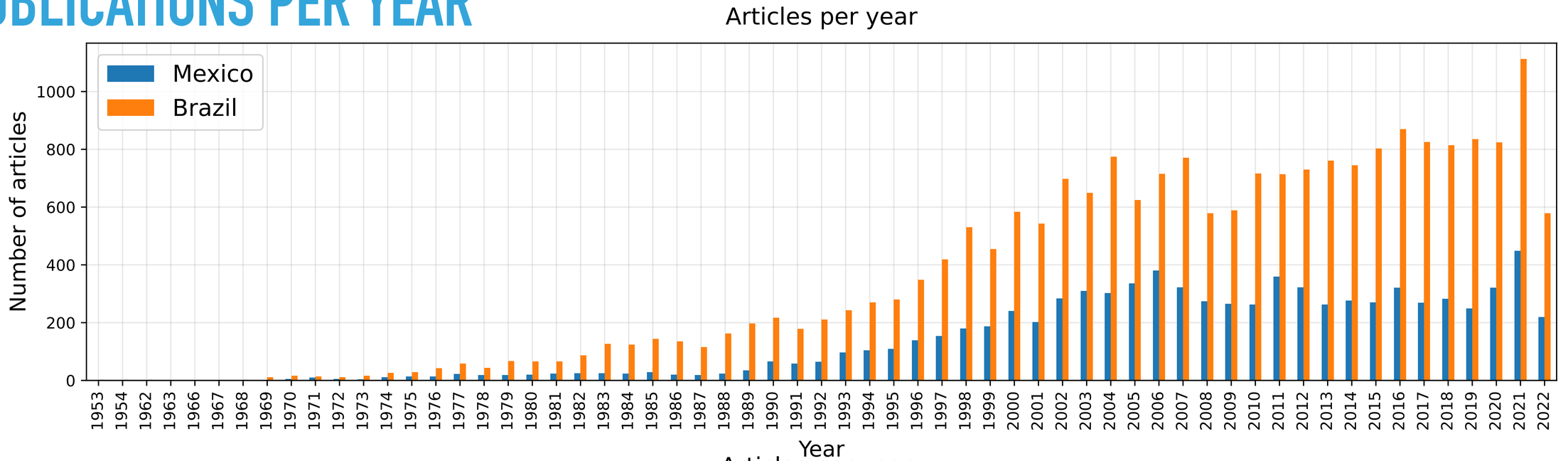
PUBLICATIONS

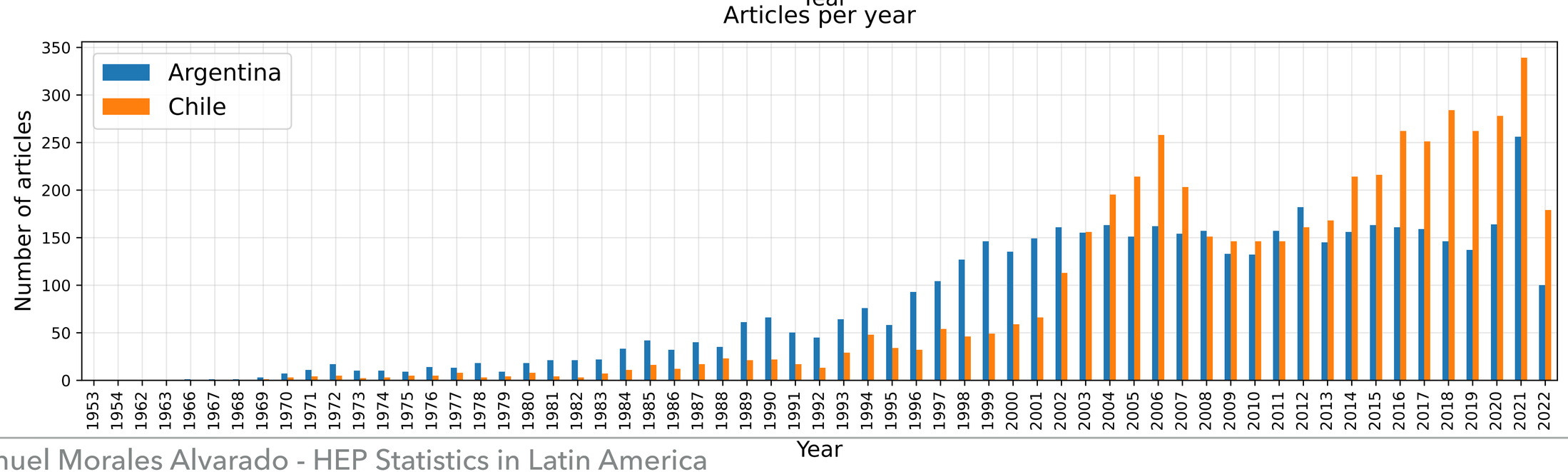
Country	Total publications	Total publications per million inhabitants
Argentina	4862	107
Bolivia	33	3
Brazil	21578	101
Chile	4976	259
Colombia	1037	20
Costa Rica	205	40
Cuba	414	37
Ecuador	112	6
Guatemala	16	1
Honduras	1	0
Mexico	8289	64
Paraguay	3	0
Peru	129	4
Uruguay	301	86
Venezuela	855	30

NORMALISED PUBLICATIONS

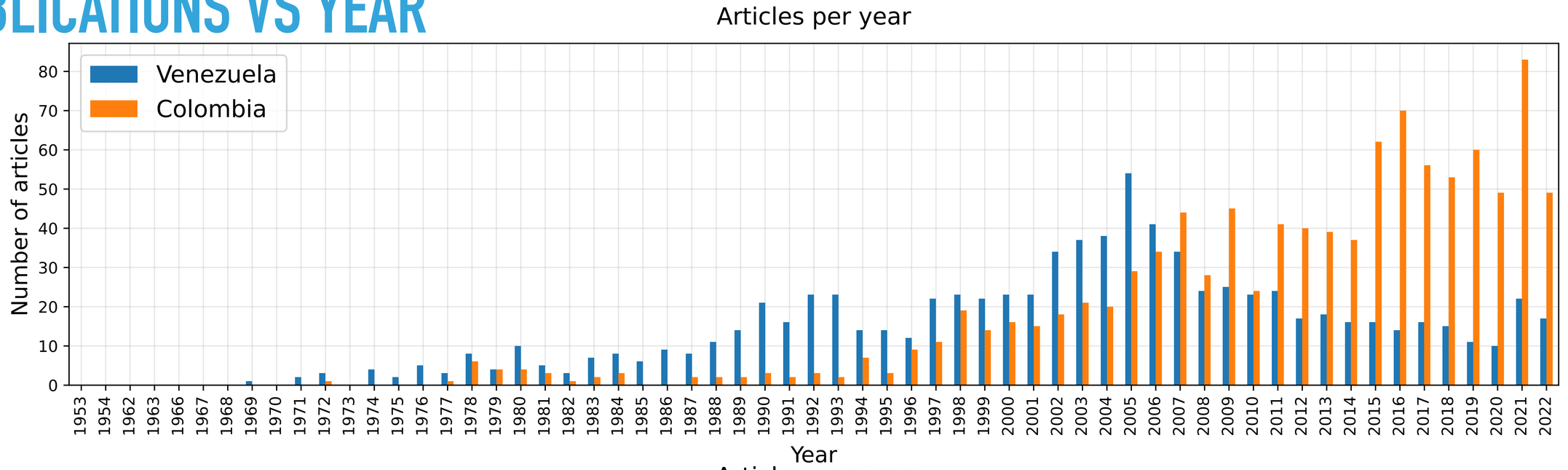


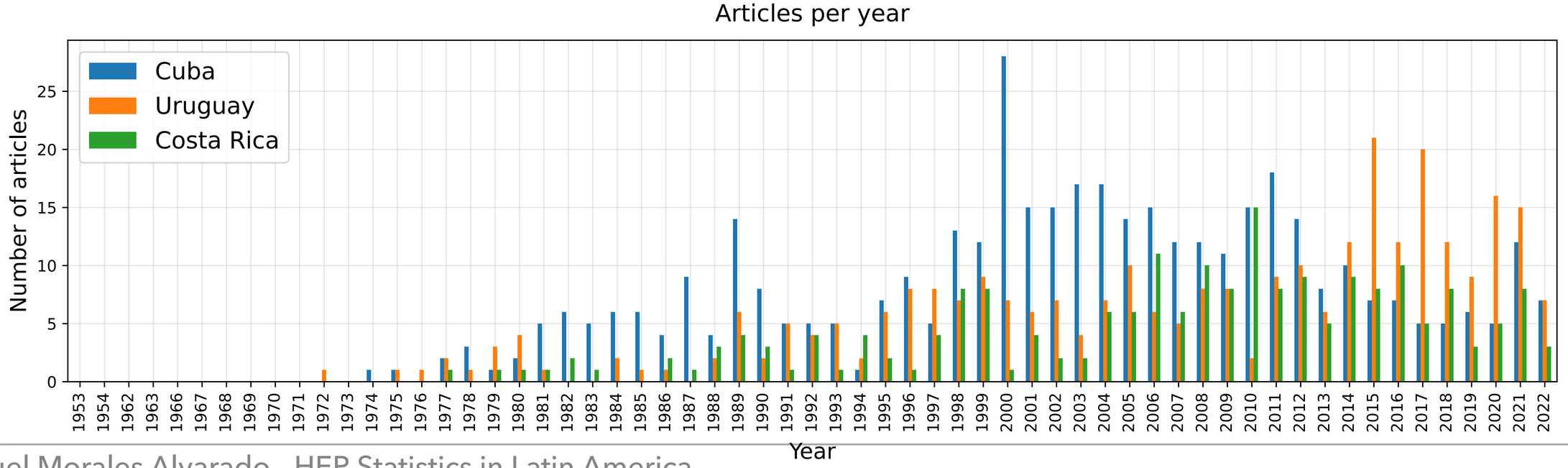
PUBLICATIONS PER YEAR



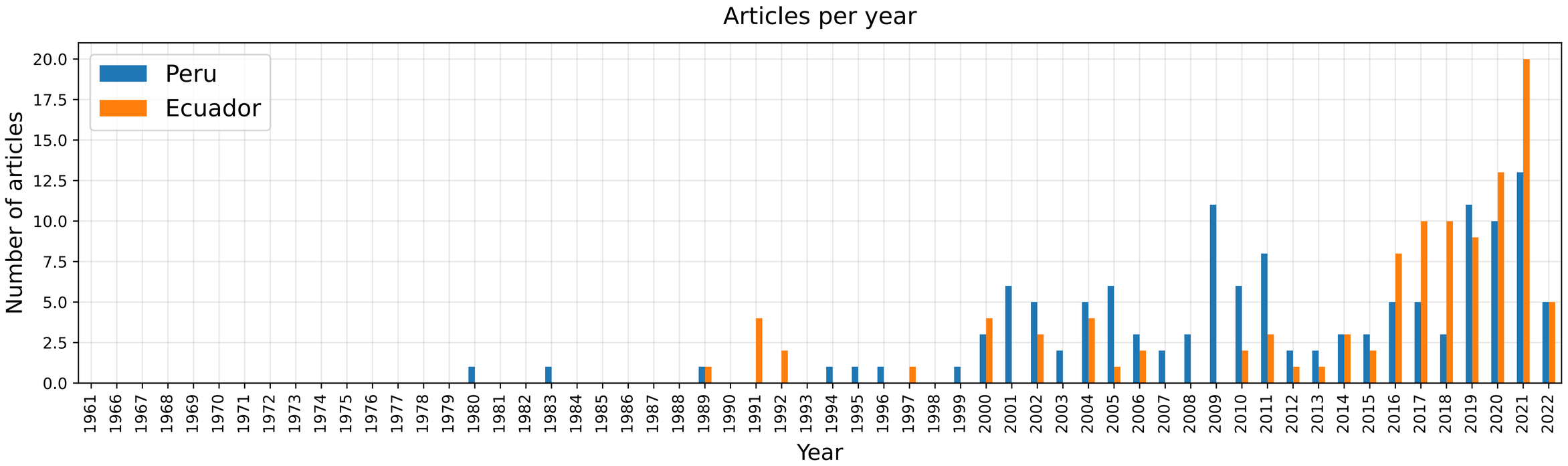


PUBLICATIONS VS YEAR





PUBLICATIONS VS YEAR

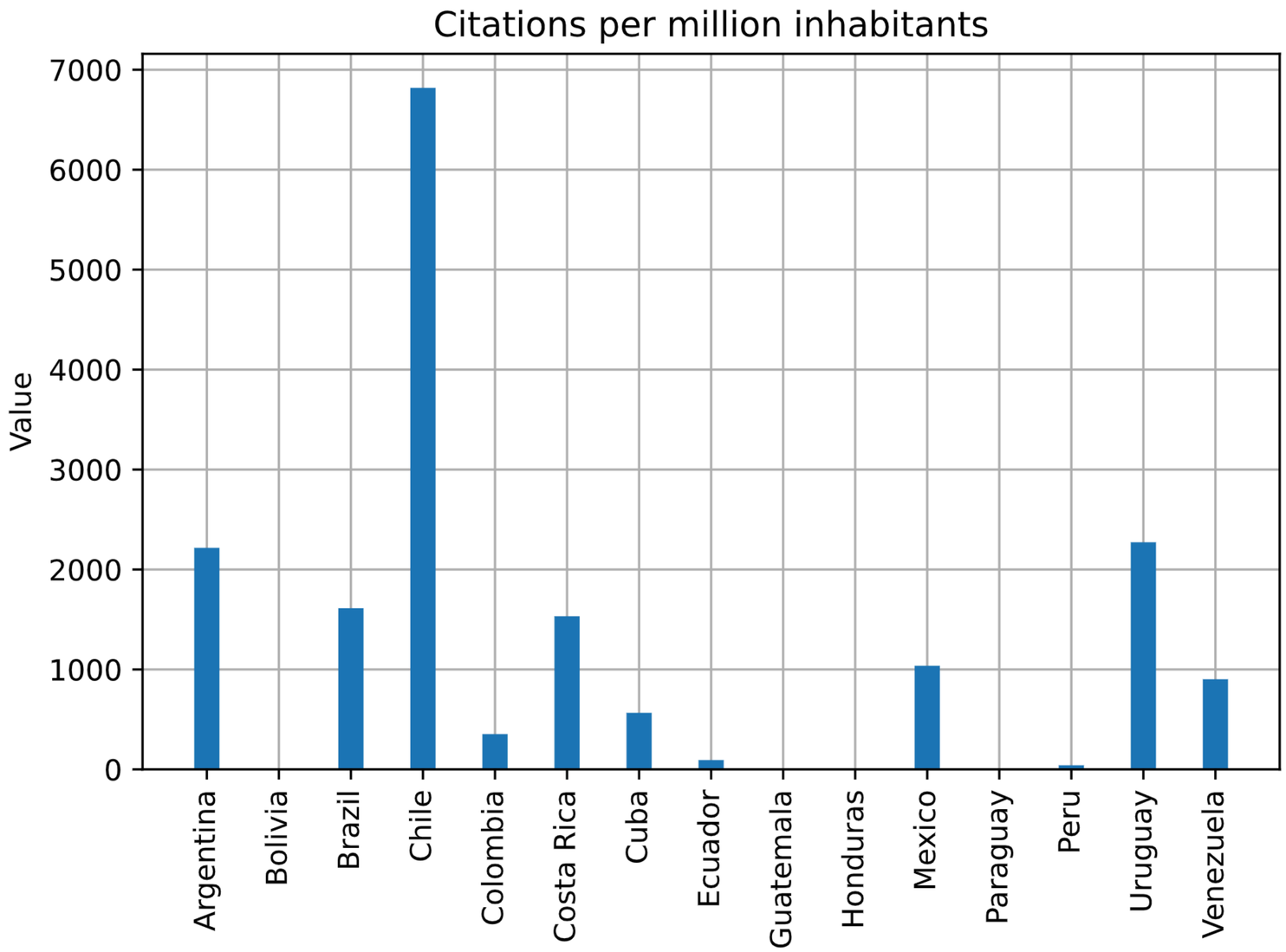


And many more countries ...

CITATIONS

Country	Total citations	Total citations per million inhabitants
Argentina	101086	2217
Bolivia	50	4
Brazil	345902	1616
Chile	130995	6818
Colombia	18118	353
Costa Rica	7876	1533
Cuba	6406	566
Ecuador	1675	94
Guatemala	109	6
Honduras	0	0
Mexico	135046	1038
Paraguay	43	6
Peru	1389	42
Uruguay	7919	2272
Venezuela	25893	902

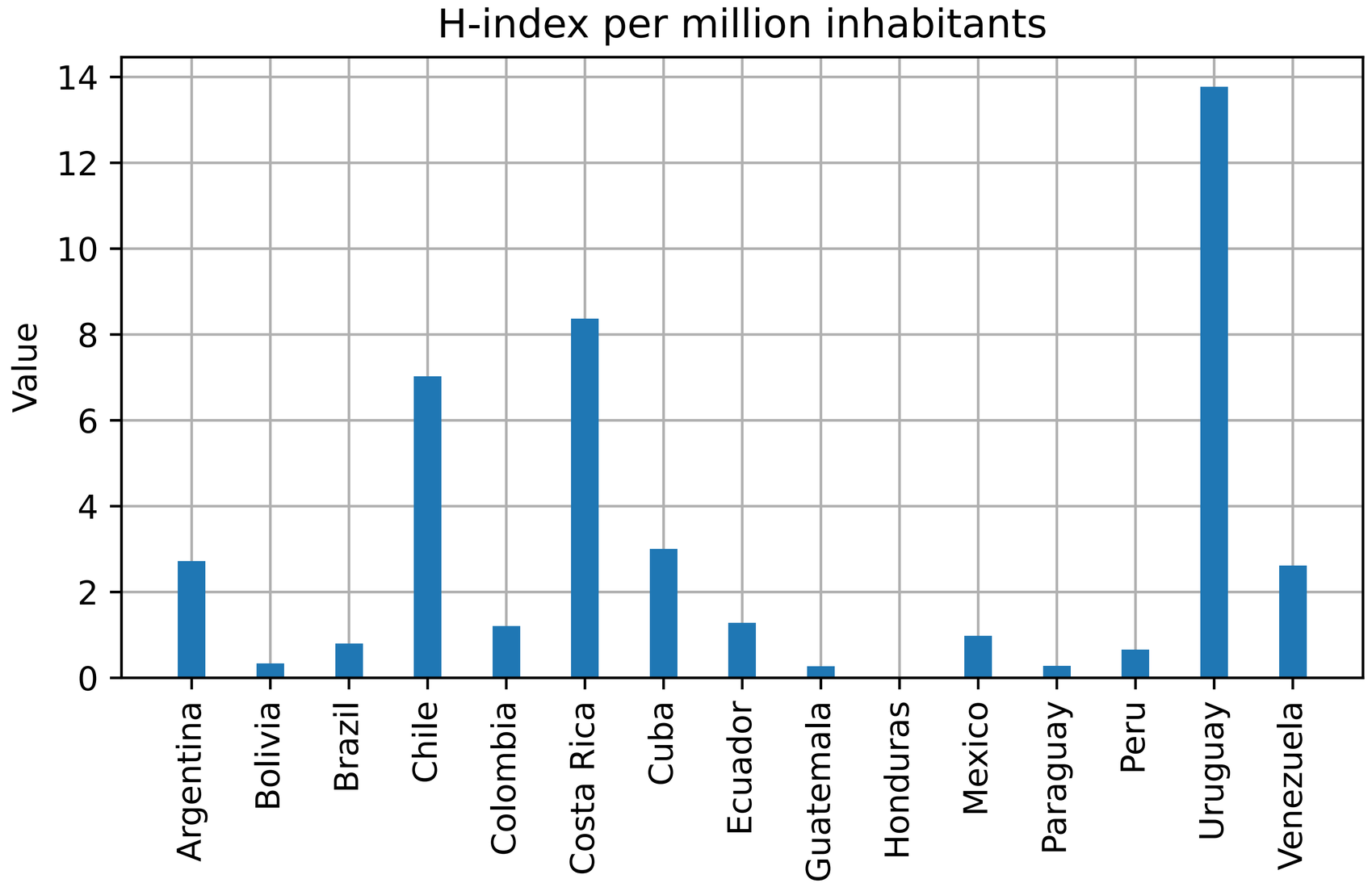
NORMALISED CITATIONS



H-INDEX

Country	h-index	h-index per million inhabitants
Argentina	124	3
Bolivia	4	0
Brazil	172	1
Chile	135	7
Colombia	62	1
Costa Rica	43	8
Cuba	34	3
Ecuador	23	1
Guatemala	5	0
Honduras	0	0
Mexico	127	1
Paraguay	2	0
Peru	22	1
Uruguay	48	14
Venezuela	75	3

NORMALISED H-INDEX



CORRELATIONS WITH GDP

Country	Percentage of GDP
Argentina	0.46
Bolivia	0.28 (2002)
Brazil	1.21
Chile	0.34
Colombia	0.32
Costa Rica	0.37 (2018)
Cuba	0.55
Ecuador	0.44 (2014)
Guatemala	0.03
Honduras	0.04 (2004)
Mexico	0.28
Paraguay	0.14
Peru	0.16
Uruguay	0.48
Venezuela	0.34 (2014)

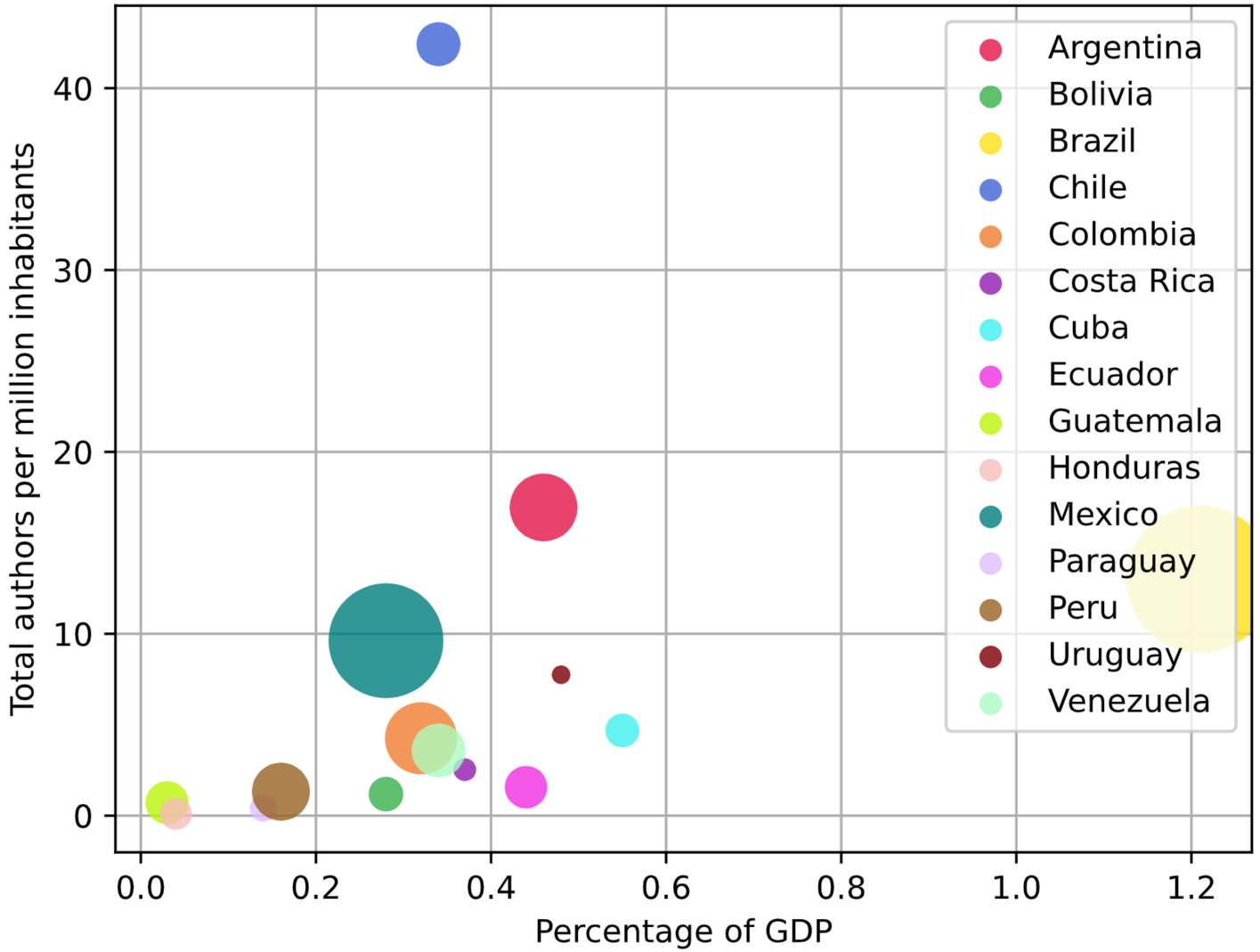
GDP = gross domestic product (invested in research and development)

Data obtained from https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS (checked in September 2022)

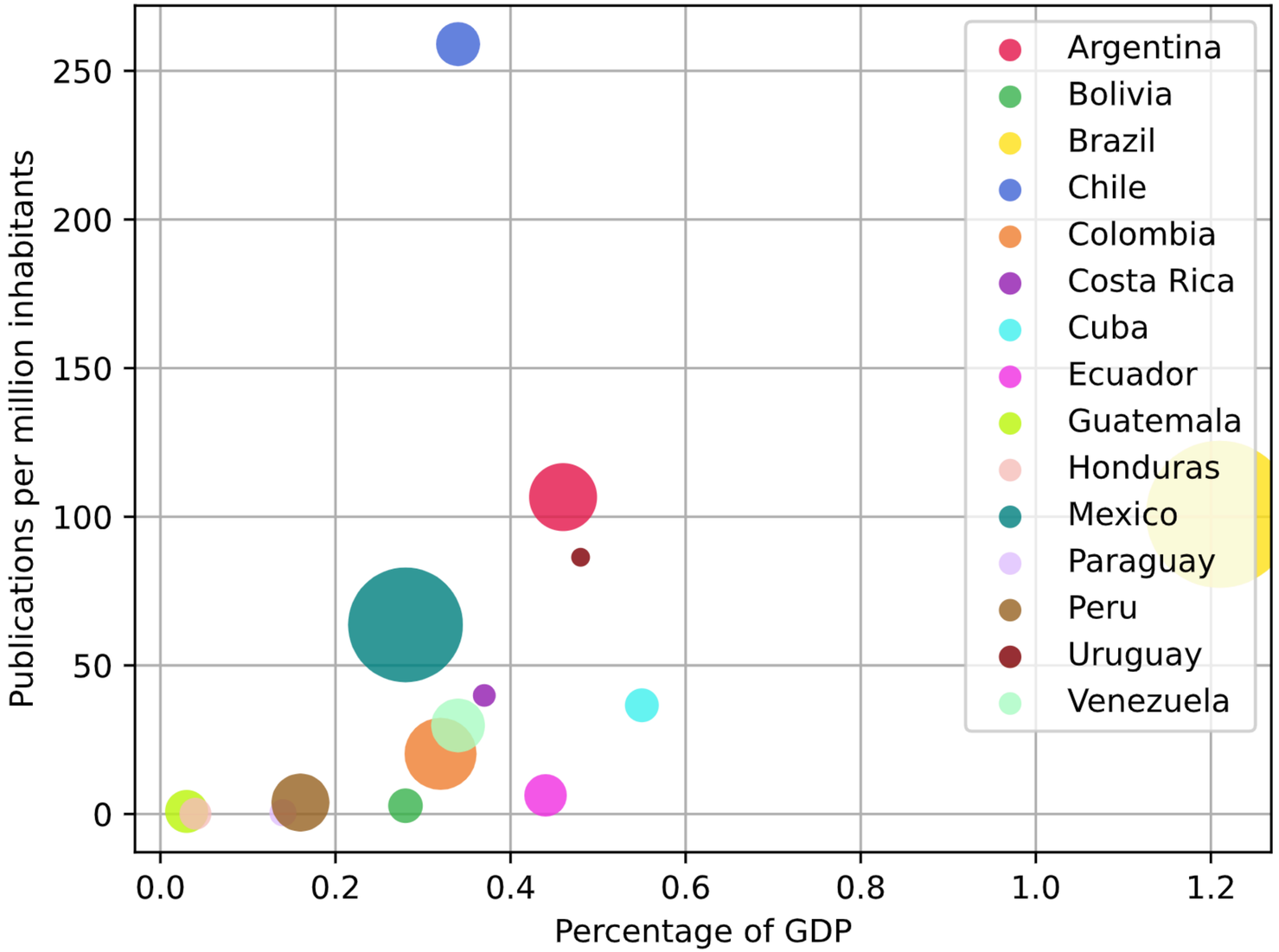
Some references (and issues):

Entity	Percentage of GDP
China	2.24
European Union	2.22
OECD	2.67
United Nations	2.33
United States	3.17

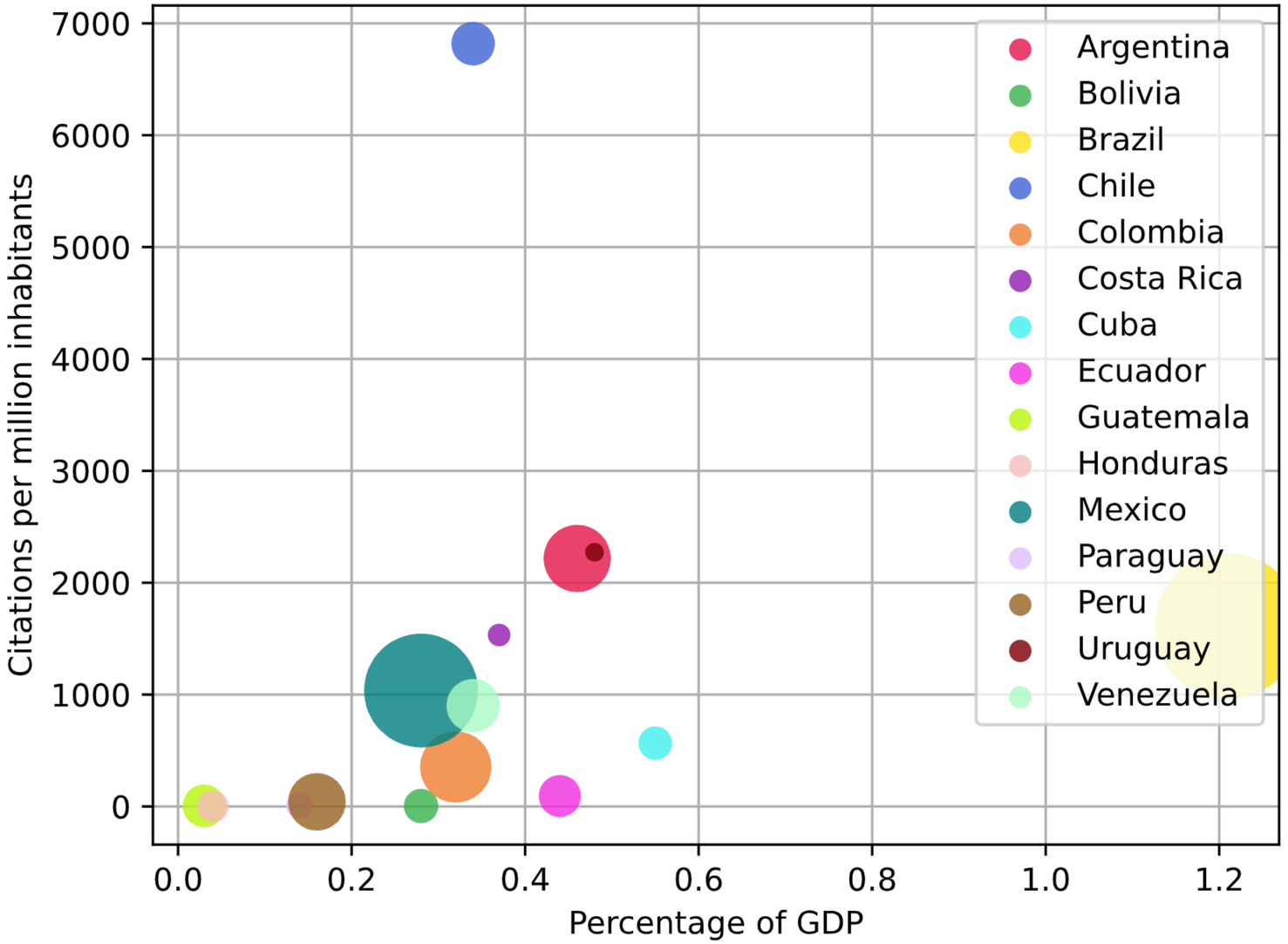
NORMALISED AUTHORS VS GDP



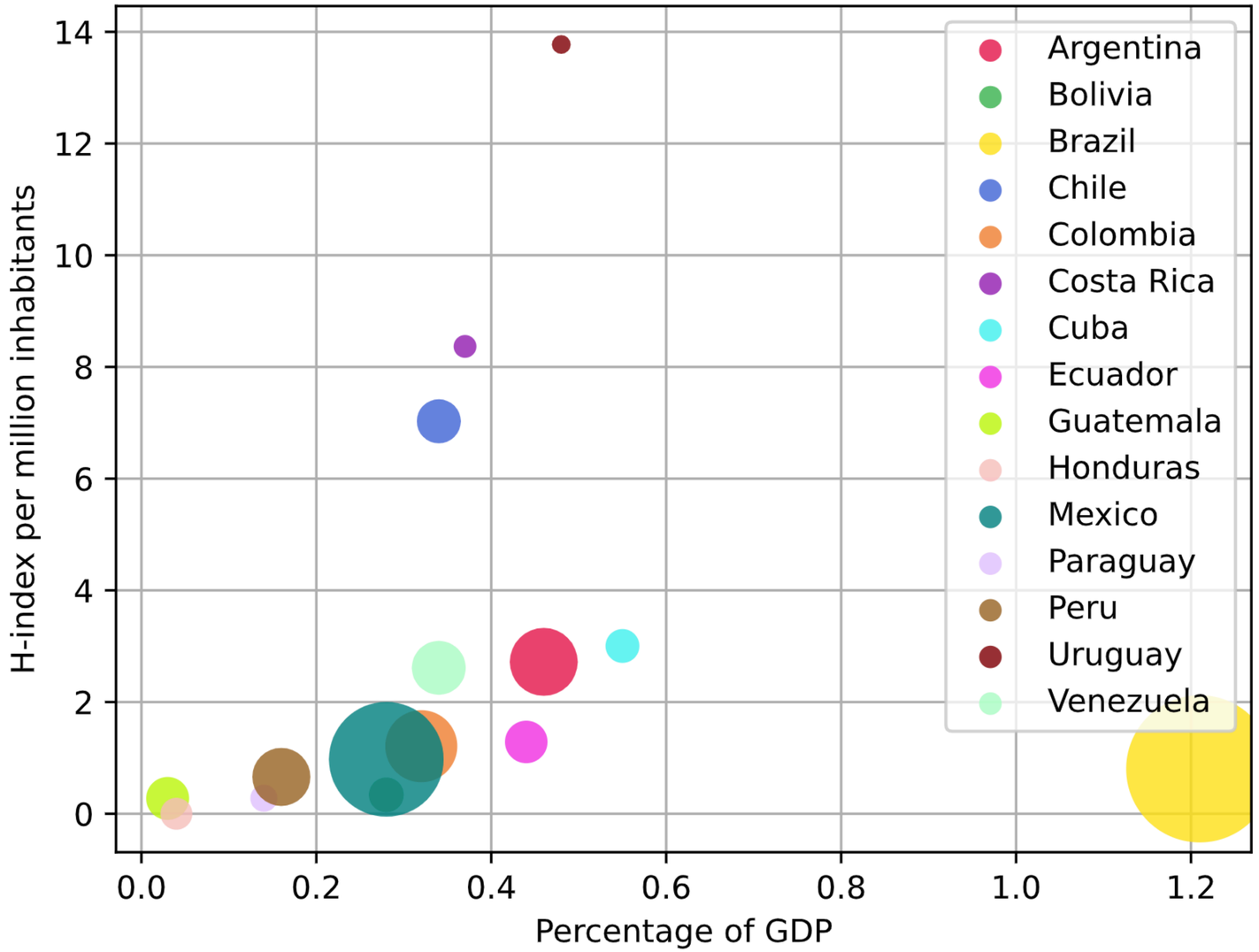
NORMALISED PUBLICATIONS AND GDP



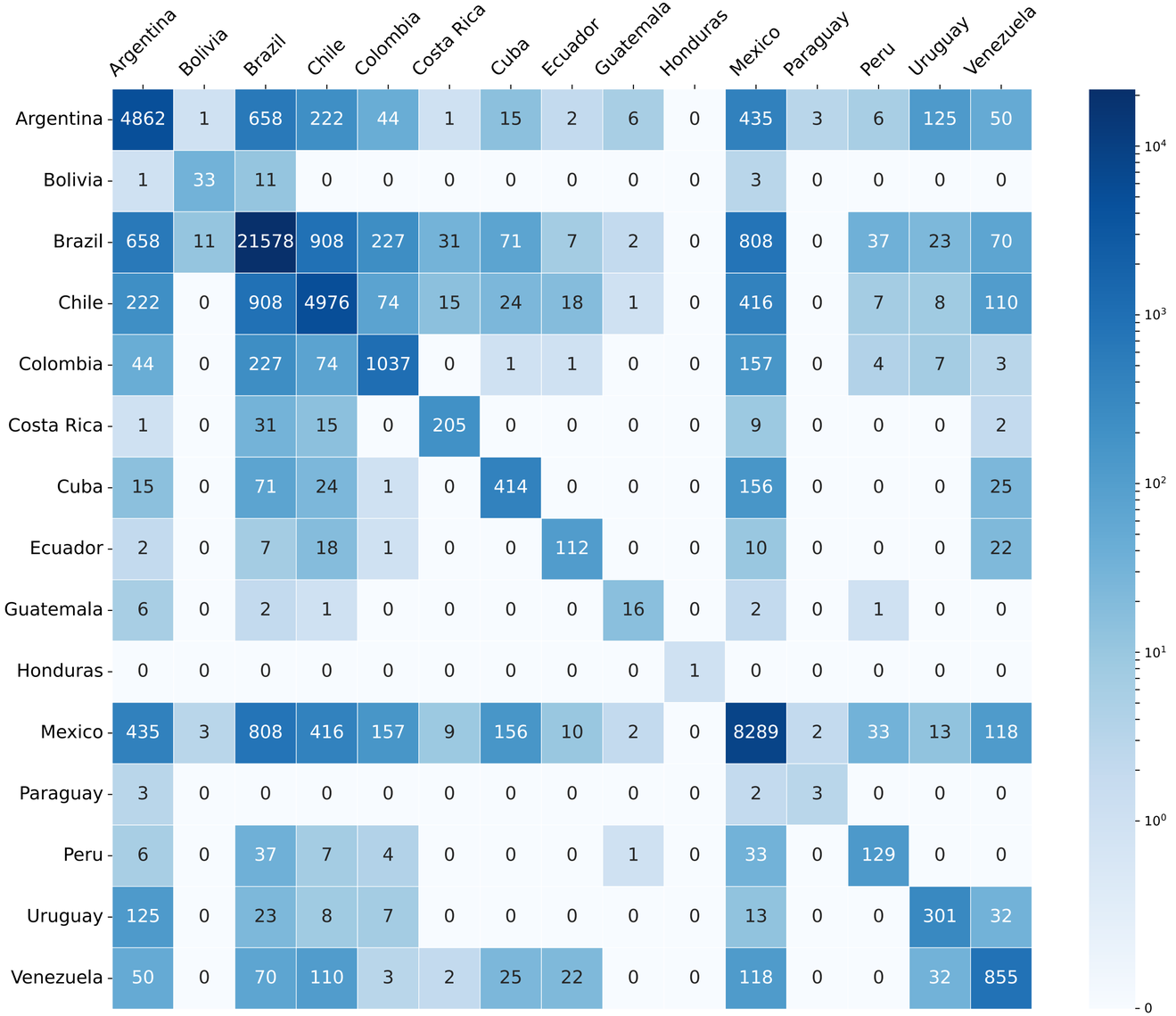
NORMALISED CITATIONS AND GDP



NORMALISED H-INDEX AND GDP

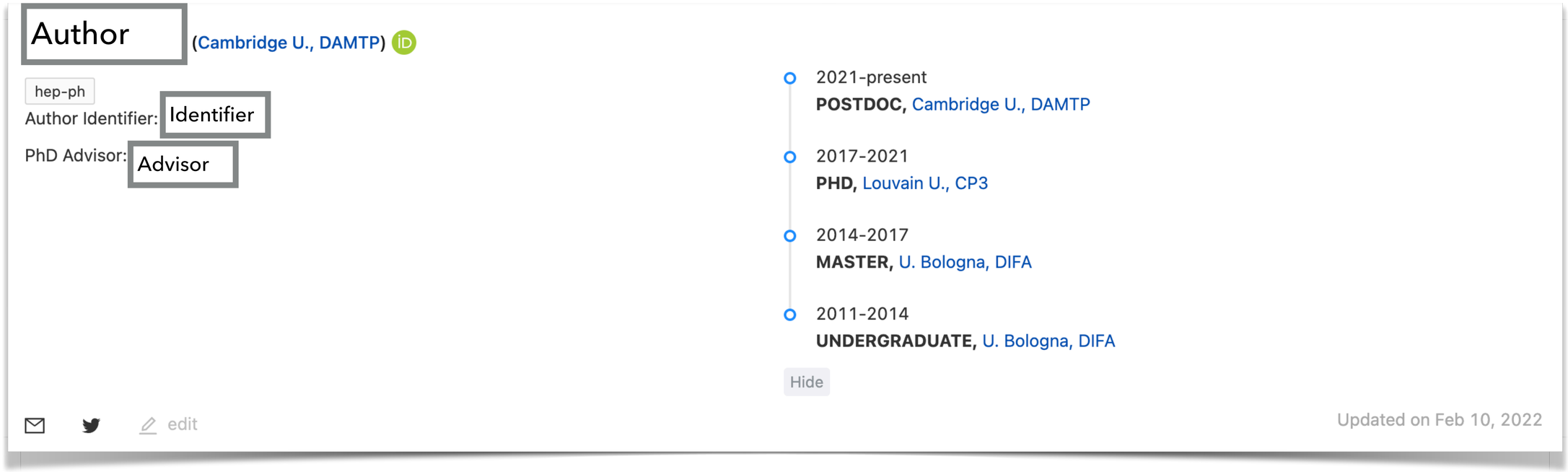


NUMBER OF COLLABORATIONS WITHIN LATIN AMERICA



CAN YOU HELP US?

Yes, by completing your INSPIRE-HEP profile



FOR THE FUTURE

- Obtain collaborations with the rest of the world
- Explore the mobility of Latin American researchers
- Extra coordination with other LAAHECAP committees (JEDI, for example)
- Extend this analysis to other areas of physics and/or science in general

CONCLUSIONS AND OUTLOOK

- We discussed HEP Statistics in Latin America

- ☑ We have a developed a database to study HEP research in Latin America
- ☑ We have developed a code to study different metrics (authors, publications, citations, etc)
- ☑ We are currently working in a extended report on the subject

- In the future:

- ☐ Study the experimental collaborations in Latin America
- ☐ Coordinate analysis with other committees in LAAHECAP
- ☐ Explore Latin American collaborations with the rest of the world
- ☐ ...

Thank you for your attention