Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782 Start Date of Project: 01/01/2019 Duration: 40 months

Deliverable 4.14 Policy API Tool

Dissemination Level	PU
Due Date of Deliverable	31/12/2019 (M12)
Actual Submission Date	11/03/2020
Work Package	WP4 - Innovations in Data Production
Task	T4.5 Social policy APIs for social surveys
Type	Other
Approval Status	
Version	V1.0
Number of Pages	p.1 – p.12

Abstract:

This document provides a description of a Social Policy API tool based on the OECD Family Policy Calculator. The document gives an overview of how the API was developed, its functionality and how it can be used. The development of API is to demonstrate how Social Policy and Contextual Data Collections can be better integrated within the wider data services offered in the EOSC.

SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.1	31/01/2020	Peer Review	Tom Emery
0.2	11/02/2020	Submission to WP Lead and Task Lead	Tom Emery
1.0	08/03/2020	Final Submission to Coordinator	Tom Emery

Author List

Organisation	Name	Contact Information
KNAW	Tom Emery	emery@nidi.nl

Time Schedule before Delivery

Next Action	Deadline	Care of
Meeting with OECD in Paris	16/10/2019	OECD
Peer Review	07/01/2020	KNAW
Submit Final Version	20/02/2020	KNAW

Executive Summary

This document sets out the rationale, design and guidelines for use of the Organization for Economic Cooperation and Development's (OECD) Family Policy Calculator Application Programming Interface (API). The API was developed and deployed as part of Task 4.5 in the Social Science and Humanities Open Cloud. The task looks to bring existing, widely used data sources into the European Open Science Cloud, increase their interoperability and integrate them within the data processing of existing Social Science Research Infrastructures. The Organization for Economic Development and Cooperations Family Policy Calculator was converted into an API and published through Plumber (a program in R) and DigitalOcean (a cloud service). This API can be queried for family policy outcomes in 35 OECD member states and provides detailed outputs regarding individuals tax position depending on their household status, income, number of children and age of the youngest child. Some examples of applications are provided, and next steps are laid out in the final section.

Abbreviations and Acronyms

API	Application Programming Interface
OECD	Organization for Economic Cooperation and Development
OECD-FPC	OECD Family Policy Calculator
EOSC	European Open Science Cloud
SSHOC	Social Science & Humanities Open Cloud
GGP	Generations & Gender Programme
SHARE	Survey for Health, Ageing and Retirement in Europe

Table of Contents

1. Introduction

Current research data management practices centre around the distribution of public use files, under limited restrictions and to a broad audience. Under this model, researchers are able to access the micro-level data file of individuals and analyse the data *in situ*. Courses in social statistics focus generally on the processing of such public use files and how they can be used to answer research questions. These practices are suitable in many instances and have many advantages. However, the processing of data in such a manner has several limitations

Firstly, some data that is useful for social research is fairly large, and potentially too large to be distributed in such a manner and efficiently loaded into popular statistical packages such as SPSS or STATA. Secondly, the analysis of such data requires significant education. Many data files are only distributed in STATA, SPSS or SAS which are statistical packages that are of limited use outside of academia. This potentially reduces the use of the data outside of traditional academic users, particularly for data scientists who might want to integrate the data into their workflow but are using software such as R or Python. Finally, distributing data in this way is ideal in cases where individual personal records are available on a finite set of individuals. In cases where the data in question is most effectively represented as a complex algorithm or complex linked data, this way of distribution becomes of limited use. Traditional public use files are poor at showing interconnections between units of analysis. Storing data in such a simple two dimensional format has been satisfactory whilst data has been cantered on random samples of individuals in which connections are not to be inferred, but with increasing links to geospatial, administrative, biomedical and social media data, this assumption is holding for less and less of the data we analyse.

Task 4.5 (Social policy APIs for social surveys) of SSHOC examines the use of API's for the distribution and maintenance of a range of social science data, with a particular focus on social policy datasets. Social policy datasets are very common in Europe and attempt to describe the policy context in a given country. The aim is then to either establish patterns of policies and change across Europe or to go further and attempt to examine how these policy patterns influence individuals' behaviour. There are a wide range of such datasets available throughout Europe, but they are almost exclusively distributed as public use files of two dimensions (i.e. an Excel sheet). This can place considerable constraints on the data that is distributed and reduces its value to researchers. This is because policies themselves are generally algorithmic in nature rather than at set values. For example, the pension policy of a country is an algorithm that determines who gets a pension and what amount. Social policy datasets have traditionally tried to represent these algorithms as static numerical values which illustrate a replacement rate, a coverage rate or an outcome value such as the percentage of the target population in poverty.

A more efficient way of distributing an algorithm is via an API. An API is a program that is hosted on a server and acts as a machine interface that returns values when provided with inputs. Social policy rules can then be coded into a common statistical language (such as R) and when queried provided with specific inputs of an individual (i.e. age, work history, marital status etc.), return outputs that describe the policy outcomes (i.e. whether they get a pension, what amount, the tax implications etc.). Such an approach allows researchers to query the API in a range of ways that can then built into the research lifecycle. For example, it might be that

someone has a dataset such as the Survey of Health, Ageing and Retirement in Europe (SHARE) and would like to know what the pension entitlements of all respondents is. These persons could simply send all respondents information to the API and receive back the complete policy outcomes for these individuals. Similarly, the API can be used to compare hypothetical situations, not based on real individuals but used as case studies to examine different ideal types of households.

In addition to the advantages for research, providing social policy information in such a way allows for simpler integration into applications and tools that draw on such data. For example, a tool could be developed that allows people to better understand their pension rights or the tax implications for having another child. Such tools currently require significant data management as part of the process as the data is generally distributed in public use files designed for research. An API would allow an application developer to simply point their application at the API and integrate the data workflow into their own application. Furthermore, in hosting policy information as an API, the policy rules can be updated and extended without the redistribution of a whole new dataset or distribution to services based on the current API.

Task 4.5 focuses on the publication of a widely used policy database as an API (SSHOC Deliverable 4.14) and demonstrates its functionality through the fielding of an innovative and interactive policy module (SSHOC Deliverable 4.15). In doing so it is hoped that many of the practical and technical advantages of hosting a social policy database as an API can be examined, tested and documented.

2. How to use the Family Support Calculator API

The Current OECD Family Support Calculator

The OECD Family Support Calculator was created by the OECD to allow easier comparisons of family support across OECD member states and facilitate policy learning across countries. The OECD has taken the rules about who is entitled to what benefits and who should pay how much tax in different countries and put them into one simple Calculator[1]. The tool, based on a modified version of the OECD's Tax and Benefit Models, allows users to choose and compare the income situation of various family types as their children age.

Tax liabilities and benefit entitlements in most countries depend on many other factors beyond those covered by the web calculator, e.g. tax-deductible expenditures, participation in employment activation programmes, unearned income and assets. Considering all these elements would reduce the scope for cross-country comparisons and add to the computational burden. The calculator therefore makes several methodological assumptions designed to ease calculations while capturing the most important characteristics of tax-benefit systems across countries. A detailed description of these methodological assumptions and other useful resources regarding the OECD tax-benefit model can be found on the project webpage[2].

Calculations include birth-related leave payments, social assistance benefits, family benefits, housing benefits and in-work benefits. Taxes and social security contributions due to earnings and benefits are also included in the calculations.

Using the Current Calculator

The current calculator is used by following the below listed steps:

1. Open the spreadsheet for the country of choice.
2. Click the "Family Support Calculator" button, which takes the user to the parameter selection window in order to choose the family type.
3. Choose the family type (single parent or couple), number of children, age of the youngest child, 1st adult's earnings level and the 2nd adult's earnings level (if applicable).
4. Click the "Calculate" button, which takes the user to the summary window providing a breakdown of earnings, benefits, taxes and net income of the selected family type.

[1] OECD Family database: the Family support calculator: http://www.oecd.org/els/soc/oecdfamilydatabasethefamilysupportcalculator.htm; 24 Feb 2020
[2] OECD Tax and Benefits project website: http://www.oecd.org/social/benefits-and-wages/; 24 Feb 2020

5. Click on the "Graph: by child age" button to study how family income evolves as the child ages and compare the income situation between various family types.

For more information, users can contact the service at the following email address: social.contact@oecd.org.

The calculator provides high quality estimates of family support, but the functionality of the calculator, as outlined in Section 1, is limited due to the fact that each individual scenario that is requested must be entered manually, and a response calculated and returned. Via an API, these values can be fed to the calculator in a structured and automated manner which allows for more expansive and meaningful comparisons. For example, if a researcher wanted to understand how family support shifted as a child ages it would be possible to feed the API a static scenario that differed only in the age of the youngest child. Entering these manually would take up to an hour. The calculator provides estimates for the first 360 months of a child's life, each of which would need to be entered into the calculator. Via the API, this could be done through a simple line of code with the results returned in seconds.

An alternative solution to this would be to distribute summary results from the calculator as a stand-alone dataset that researchers could analyse independently. However, this is also inefficient. The calculator covers over 1,197,100 separate scenarios and for each scenario estimates seven separate outputs. The resulting dataset for a complete output of the calculator is therefore in excess of 5GB which is a large dataset, especially when most researchers simply extract the information they need and discard the rest. In such instances, it is far easier to make social policy information available via an API.

Using the OECD Family Support Calculator API

Given that the data underlying the API is the same as the standard calculator hosted on the OECD website, it is substantively the same in operation. From a technical perspective, the OECD Family Policy API is provisionally hosted on the GGP server at the Dutch Academy of Arts and Science and published using Plumber[3] in R. To submit requests to the API, a user simply submits a Http request to:

```
ggp.nidi.nl:8080/a="XX"&b=X&c=X&d=X&e=X&f=X
```

where a-e are the specification of parameters to be used in the calculator:

a) Two letter code of country (ISO 3166-1 alpha-2)
b) Number of Children (0, 1, 2, 3, 4)
c) Income of Respondent as a % of Median Income (50, 100, 150)
d) Income of Partner as a % of Median Income (50, 100, 150, NA)

[3] Plumber website: https://www.rplumber.io/; 24 Feb 2020

e) Age of youngest child in months (-12 to 324)
f) Couple Status (1 = Single; 2 = 2 earner household)

For example, if a researcher would like to submit a request for the outputs of a couple earning median income in the United States with one child who is six months old, they would submit the following request:

```
ggp.nidi.nl:8080/a="US"&b=1&c=100&d=100&e=6&f=2
```

A researcher could then extract the results and submit a new request which extracts the results for the couple when their child is one month older to see how the policy position changes over time. The machine interface nature of the API allows this to be executed rapidly and systematically to build complex and flexible policy simulations. Submitting the Http request with valid values returns a vector in JSON format which lists the following in order, expressed in the country's local currency;

1- Total Household Earnings
2- Housing Benefits (assuming costs of 20% of Average Wage)
3- Family Benefits
4- Social Assistance Benefits
5- In-work Benefits
6- Unemployment Benefits
7- Paternity Leave Benefits
8- Parental Leave Payment
9- Income Tax (& Tax Credits)
10- Social Contributions
11- Net Income.

3. Conclusions and next steps

The OECD Family Support Calculator is a powerful source of data on social policy support for families across the OECD. Its publishing as an API is helpful in making the data more accessible and interoperable. The data from the calculator can be readily accessed for the first time and the algorithm queried in a machine actionable manner. The OECD Family Support Calculator is only one of a wide range of social policy datasets that are maintained across Europe and beyond. Many of these would operate more effectively and efficiently if an API was made available, making the integration of social policy information into social analyses simpler and more efficient. Several examples of such datasets and data systems include Euromod (Tax-benefit microsimulation model for the European Union[4], SPIN[5], The OECD Tax/Benefit Model[6], The Mutual Information System in Social Protection (MISSOC)[7], The ILO Social Security Database Programs[8] and many others.

However, within these datasets the data is either stored as text-based descriptions of policy prescriptions or as quantitative indicators of policy coverage for ideal household types. This is because such datasets either seek to distribute a PDF describing the policies or a dataset which can be downloaded and directly used in research. Publishing these systems as an API allows for the policy conditions to be coded into an algorithm, be centrally maintained and updated whilst allowing users to work flexibly and efficiently with the algorithm directly. To support this, task 4.5 will seek to create a short webinar which demonstrates how existing social policy databases can be published as an API and catalogued within the EOSC. There is a wide range of such databases that exist and would be of high value to a big number of researchers and infrastructures. The production and maintenance process of such API's is minimal given that documentation of the data and algorithms generally already exists. Before this is conducted, standards of documentation and a public cataloguing process should be identified.

Furthermore, the publishing of the calculator as an API will allow for the integration in online surveys such as the Generations and Gender Programme. To demonstrate this potential, a module will be run in an online version of the Generations and Gender Survey which will draw on the OECD Family Support Calculator as part of Deliverable 4.15. In this module, respondents will be asked about their current family support position and how it might be affected by changes in their circumstances. The API will allow for these responses to be validated against the calculator in real time, and for follow up probing questions to understand individual's

[4] Euromod (Tax-benefit microsimulation model for the European Union) website: https://www.euromod.ac.uk/; 24 Feb 2020
[5] Social Policy Indicators website: https://www.spin.su.se/; 24 Feb 2020
[6] OECD webpage: https://www.oecd.org/social/benefits-and-wages; 24 Feb 2020
[7] The Mutual Information System in Social Protection website: https://www.missoc.org/missoc-database/; 24 Feb 2020
[8] The ILO Social Security Database Programs website: http://www.ilo.org/sesame/ifpses.socialdatabase; 24 Feb 2020

perception of family support policy. This has the potential to generate significant insights in the field of social policy and how social policy effects peoples' behaviour, such as their plans to have another child.