SIA

SOCIAL
INVESTMENT
AGENCY
ORANGA TANGATA

Investing in what works for better lives

Social Investment Analytical Layer Code User Guide

Guidelines for creating events tables for social investment analysis*

Version 1: December 2017
SIA-2017-0248

New Zealand Government

Creative Commons Licence

Liability

Citation

Intended audience

This guide is written for analysts and policy people to support data analytics. However we have adopted a plain English style of language to make the guide as accessible as possible.

Resources, tools and guides

The SIA is developing a range of tools, products and guidance to enable agencies to develop their social investment approaches, and analyse and measure the impact and effectiveness of the services they're delivering.

* By 'social investment analysis' we mean analysis of anonymised, integrated data with the objective of measuring the impact of government services on people's lives.

Analysing data for social investment

Analytic methods are vital to the social investment approach. We need to be able to see the effectiveness of social services across the full spectrum of outcomes known to the social sector, and to be able to test whether or not the initiatives we deliver are benefiting the New Zealanders who are receiving them.

Statistics New Zealand provides a world-leading resource for social investment analytics: the Integrated Data Infrastructure (IDI). It is a research database containing multiple datasets collected by government agencies and non-governmental organisations.

Using this integrated data resource, analysts are able to generate insights which help us understand what works, for whom, at what cost.

Analytical tools to simplify analysis

The datasets available in the IDI are not structured in a standard accessible way. To make it easier for analysts to make use of this data, the SIA has created tools which simplify its use. The focus of this guide is the Social Investment Analytical Layer (SIAL). The SIAL is a piece of code which creates events-based tables from a selection of the data available in the IDI.

The SIAL was developed to be used in conjunction with a suite of analytical tools produced by the SIA:
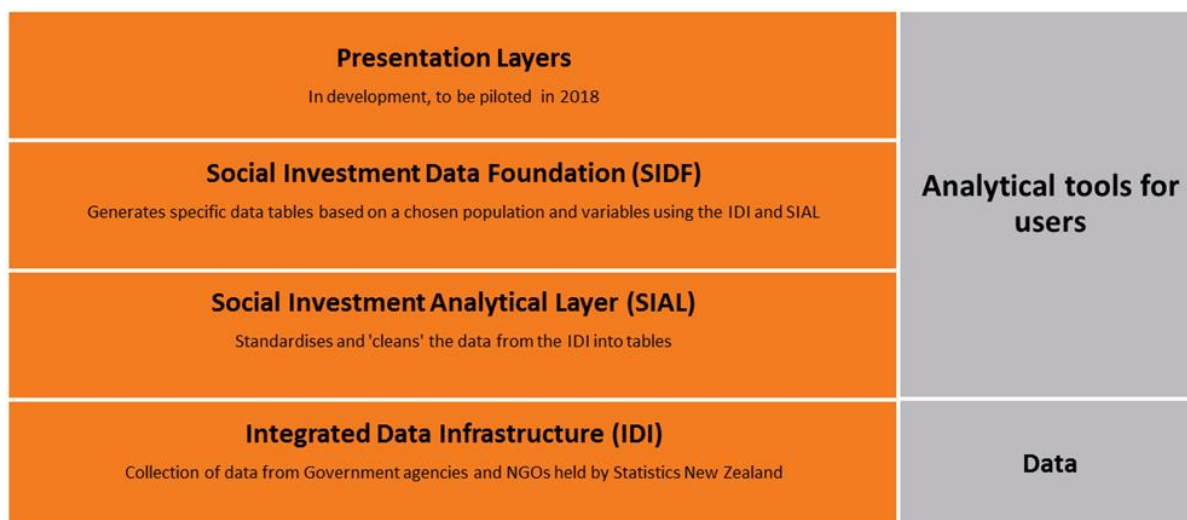
The tables generated by the SIAL code are one input into the Social Investment Data Foundation (SIDF) code. The SIDF builds on the SIAL code to produce a dataset that is ready for analysis. The SIDF user guide is available on the SIA website. We recommend that all analysts working with IDI data make use of these tools to save time and effort, and avoid duplication of work.

This guide explains how to apply the SIAL code to enable analysis of IDI data. Although the SIAL code is publically available, you will need access to a Statistics New Zealand data lab to apply it to IDI data. You can learn more about how to apply for IDI access in the SIA's Beginner's Guide to the IDI.

The Social Investment Analytical Layer

The SIAL code cleans and structures each dataset, turning the unstructured data available in the IDI into standardised tables.

Standardised tables make analysis faster and more consistent

The SIAL code structures the administrative data into social sector service events-based tables, and documents business rules to the formatted data. These standardised tables allow users to analyse the social sector services people receive in a more user-friendly way. This format encourages a consistency in the definitions of variables, makes analysis faster, and enables greater understanding of the data.

Social sector service event tables

A social sector service event table is a table which contains all occurrences of a particular service delivered to individuals. Each service is considered an event, for example: a benefit received, a corrections sentence served, a hospital visit, or time in school, are all considered social sector service events.

Table 1 shows all of the social sector service events tables available in the SIAL, and the date they are available from. Figure 1 provides the same information as a timeline diagram.

Table 1: Social sector service events tables available in the SIAL

Agency	Subject area	Description	Available from
MSD	T1	Main benefits	1993
MSD	T2	Supplementary benefits	1993
MSD	T3	Hardship payments	1993
MOE	INT	Student interventions (e.g. truancy)	1997*
MOE	ENR	Primary and secondary school enrolments	2007
MOE	ECE	Early childhood education	2007*
MOE	TER	Tertiary enrolments	1994
MOE	IND	Industry training	2003
IR	W&S	Wages and salary	1999
IR	PEN	Pension payments	1999
IR	CLM	ACC weekly compensation	1999
IR	STU	Student allowance	1999

Agency	Subject area	Description	Available from
IR	BEN	Benefit payments	1999
IR	PPL	Paid parental leave	2002
ACC	INJ	ACC injuries	1994
MOJ	CHA	Court charges	2004
COR	SAR	Corrections sentencing and remand	1970
MOH	CHR	Chronic conditions	1988
MOH	PFD	Publically funded hospital discharges	1989
MOH	CAN	Cancer registrations	1995
MOH	GMS	General medical subsidy (after hours and unenrolled patients)	2002
MOH	LAB	Lab tests	2003
MOH	PHA	Pharmaceutical dispensing	2005
MOH	NPA	National non-admitted patients (outpatients and emergency department)	2007
MOH	PRI	PRIMHD - mental health collection	2008*
MOH	B4S	Before school health checks	2011
HNZ	REG	HNZ register waitlist events	2000
CYF	CEC	Child Youth and Family events	1992*
CYF	ABE	CYF abuse findings	1994*
POL	VIC	Police records of victimisations	2014
POL	OFF	Police records of offenders	2009
MIX	SLF	Self-harm service events. Mixed agency table.	1989
MIX	MOR	Mortality service events. Mixed agency table.	1840

* Data is available from this date although better quality data is available in more recent updates.

Figure 1: Social sector service events tables available in the SIAL
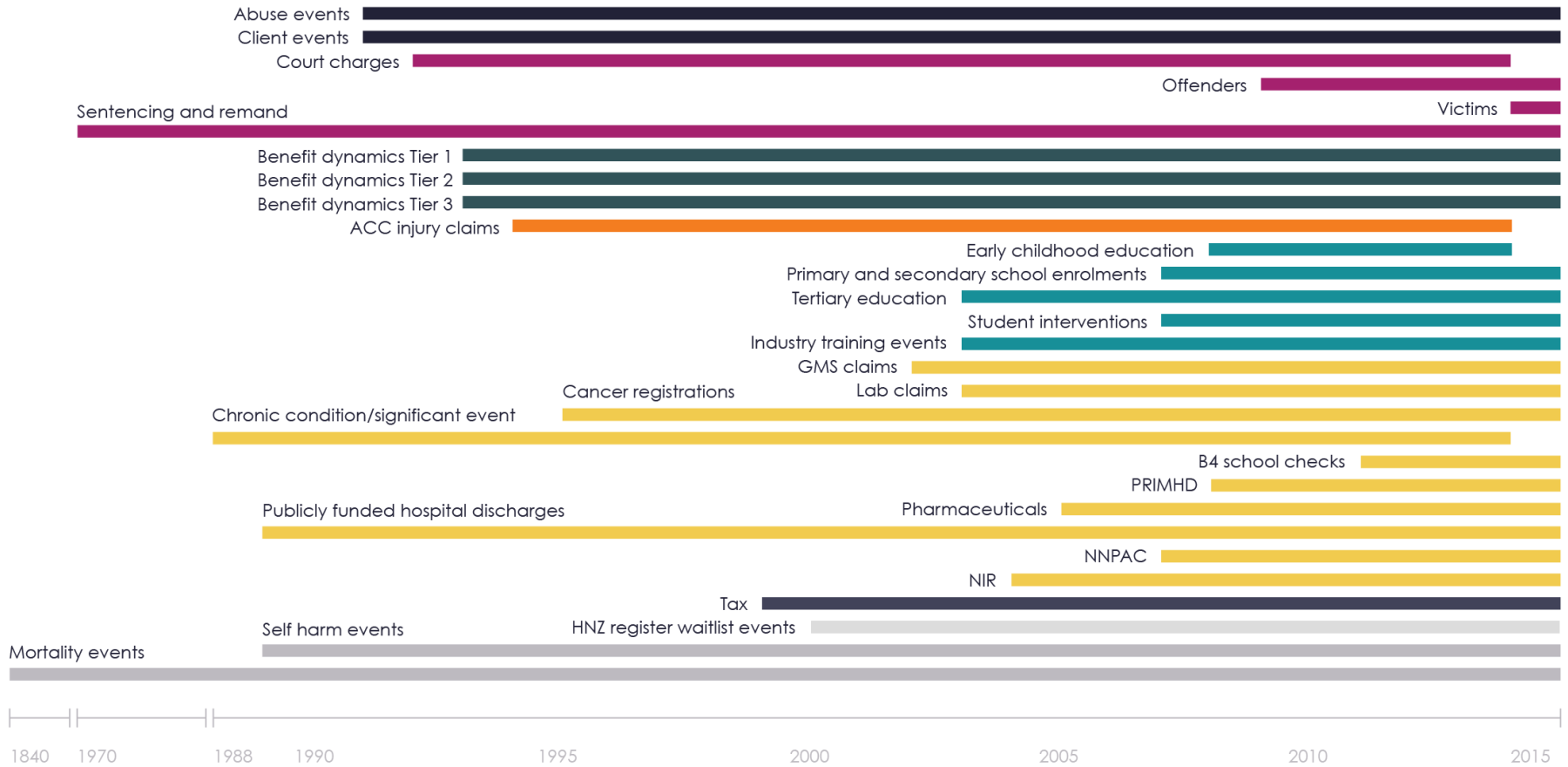
| | CYF | Justice | MSD | ACC | Education | Health | IR | Housing | Mixed agency |

Social sector service event table structure

The social sector service event tables that make up the SIAL all have the same structure.

They are structured as one row per person, per event spell, and have the following variables:

- Unique identifier (snz_uid)

- Department (department)

- Data mart (datamart)

- Subject area (subject_area)

- Start date (start_date)

- End date (end_date)

- Event type(s) (event_type – event_type_n)

- Cost(s) (cost – cost_n).

Table 2: Example of a service event table using synthetic data

snz_uid	department	datamart	subject _area	start_date	end_date	event _type	cost
1234	MSD	BEN	T1	01/01/2013	31/12/2014	JSS	$10,987.65
9876	MSD	BEN	T1	03/03/2014	12/12/2014	YPP	$1,234.56
1234	MSD	BEN	T1	30/06/2015	30/09/2015	JSS	$1,234.56

Accessing and using the SIAL code

The SIAL code is hosted on [GitHub](#) – a dedicated online hosting service where the open source community share code.

While the SIAL code is publically available, you will need access to a data lab to apply this code to the IDI. You can read how to apply for access in the Social Investment Agency's Beginners Guide to the IDI.

The code needed to build the SIAL tables, data dictionaries, and business rules can be viewed and downloaded at:

https://github.com/nz-social-investment-unit/social_investment_analytical_layer

To use the code:

- Download the zip file from GitHub

- Email access2microdata@stats.govt.nz (the Microdata team at Statistics NZ) asking them to put the code into your own project for your own use.

The SQL code generates views (a convenient way to save space and ensure data is up to date) while the SAS code generates tables.[1]

Can I contribute a new events table that does not exist to the SIAL code?

Yes! The Social Investment Agency encourages collaboration and sharing.

To create code to generate a new SIAL table (or tables) these are the steps:

1. Your events table needs to have the correct format
2. You need to document your business rules and data dictionary[2]
3. The organisation the data is collected from needs to review your business rules and code (although code can be made accessible while business quality assurance is on-going).

Service event table format

Table 3: Format of service events tables in the SIAL

Variable Name	Description	Compulsory	Values
snz_uid	Unique identifier	Yes	Numeric identifier
department	Organisation	Yes	Character, 3 E.g. MOJ = Ministry of Justice
datamart	Data mart	Yes	Character, 3 E.g. COU = Courts
subject_area	Subject of events	Yes	Character, 3 E.g. CAS = Cases
start_date	Start date of event	Yes	Datetime
end_date	End date of event	Yes	Datetime
event_type	Description of event type	Yes	Character, < 10

---

[1] **SQL** (Structured Query Language) – a domain-specific language used in programming.

SAS (Statistical Analysis System) – a data processing and statistical analysis language

[2] **Business rule** – a rule that defines or constrains some aspect of business and always resolves to either true or false. Business rules are intended to assert business structure or to control or influence the behaviour of the business.

Data dictionary – a set of information describing the contents, format, and structure of a database and the relationship between its elements, used to control access to and manipulation of the database.

Variable Name	Description	Compulsory	Values
			E.g. Type of court 'Adult' or 'Youth'
event_type_n	Additional description of event type. n starts from 2; event_type, event_type_2, event_type_3, etc.	No	Character, < 10 E.g. 'PVN' = Proven; 'UNP' = Unproven
cost	Direct event cost as a lump sum for the length of the event.	No	Numeric
cost_n	Other non-direct event costs as a lump sum for the length of the event.. n starts from 2; cost, cost_2, cost_3 etc.	No	Numeric

Business rules and data dictionary

You need to supply Statistics New Zealand with a detailed account of changes made to the data tables and steps taken in your research for them to consider the release of your results.

Document any business rules you applied when restructuring the administrative data table into an events-based table, and clearly state how you derived any variables, for example, costs. If you didn't need to restructure the table or derive any new variables that's OK, please just state this and provide the data dictionary.

Review

The organisation the data is collected and supplied by needs to review your business rules and code. Once you have completed these (or your business quality assurance is in progress) please contact info@sia.govt.nz to add your events table to the SIAL.