CA2SIG Standards WG

Session 1

Christiaan Pauw

4/13/2021

Introductions and welcome

Proposed workplan for 2021

Objectives

Why DLT ?

Why Standardisation ?

What will the working group deliver and who will use it?

There are other bodies that create the types of standards mentioned above. This working group will be more concerned with how such standards themselves and compliance to these standards are encoded and captured in a distrubuted ledger and how such representations can be used.

Four phases

A. Conceptualisation of a DLT-enabled standardisation architecture to streamline Climate Action & Accounting.

B. Implementing the result of phase A in three different paradigms:

- as a relational database,
- as a Radix-Merkle tree (as is used by Hyperledger Sawtooth), and
- as a token-and-smart-contract network (possibly DAML).

C. Testing applicability to real-life use cases.

D. Reviewing the results from phase C and revisiting phases A-C as necessary.

Phase A

Why not just make a digital representation of the existing standard?

- Because there are many standards
- They have different aims and uses
  - Voluntary or Compliance
  - Narrow climate focus or broader social goals
  - Time frame (short-lived climate pollutants [SLCPs] vs. long term effects)
  - Align MRV for mitigation and adaptation (very early days for adaptation standards)
- Align to broader SDG context

Example: DeclareDesign

DeclareDesign.org is an example of a elegant solution for a pervasive problem (in this case how to describe a research design before starting the study) based on a process where:

- a conceptual framework was created,
- the applicability was tested / extended to classic use cases
- software implementations and tools were created to easily apply the framework

Additional example: Zotero

Example: DeclareDesign (continued)

DeclareDesign: : CHEAT SHEET

Model

What is your model of the world, including how outcomes respond to interventions in the world?

Population

Define the size of the population, hierarchical structure (if any), and background variables.

Simple dataset with no background variables

```
pop <- declare_population(N = 100)
pop()
```

Simple dataset with background variables

```
declare_population(N = 100,
                   X = rnorm(N))
```

Two-level dataset

```
declare_population(
  schools =
    add_level(N = 10,
              funding = rnorm(N)),
  students =
    add_level(N = 100,
              scores = rnorm(N))
)
```

Outcomes

Outcomes that depend on a treatment (Z)

Using a formula

```
declare_potential_outcomes(
  Y ~ .5 * Z + rnorm(N))
```

As separate variables

```
declare_potential_outcomes(
  Y_Z_0 = rnorm(N),
  Y_Z_1 = Y_Z_0 + .5)
```

Outcomes that do not depend on treatment

```
declare_potential_outcomes(
  Y = rnorm(N))
```

Inquiry

What is the research question you want to answer?

Causal inquiries

```
declare_estimand(
  ATE = mean(Y_Z_1 - Y_Z_0))
```

Descriptive inquiries

```
declare_estimand(
  Y_median = median(Y))
```

Conditional estimands

```
declare_estimand(
  LATE = mean(Y_Z_1 - Y_Z_0),
  subset = complier == TRUE)
```

Data Strategy

How will you generate data to answer your inquiry?

Sampling

```
declare_sampling(n = 100)
```

```
declare_sampling(
  strata_n = 20,
  strata = urban_area)
```

Treatment assignment

```
declare_assignment(m = 100)
```

```
declare_assignment(
  clusters = villages,
  m = 10)
```

Answer Strategy

How will you generate an answer to your inquiry?

OLS with robust standard errors

```
declare_estimator(
  Y ~ Z, model = lm_robust)
```

2SLS instrumental variables regression with robust SEs

```
declare_estimator(
  Y ~ D | Z, model = iv_robust)
```

Difference-in-means

```
declare_estimator(
  Y ~ Z,
  model = difference_in_means)
```

DeclareDesign is a software implementation of the MIDA framework, according to which research designs have a **M**odel of the world, an **I**nquiry about that model, a **D**ata strategy that generates information about the world, and an **A**nswer strategy that uses data to make a guess about the **I**nquiry. Declared designs can be "diagnosed" to calculate the properties of the design such as power and bias using Monte Carlo simulation.

All `declare_*` functions return *functions*. Most functions take a data.frame and return a data.frame.

Design Declaration

Put together all the steps into a declared design using the + operator

```
design <-
  declare_population(N = 200, X = rnorm(N)) +
  declare_potential_outcomes(Y ~ .5 * Z + X) +
  declare_estimand(ATE = mean(Y_Z_1 - Y_Z_0)) +
  declare_sampling(n = 100) +
  declare_assignment(m = 50) +
  declare_estimator(Y ~ Z, model = lm_robust)
```

```
draw_data(design)
draw_estimates(design)
get_estimates(design, data = real_data)
draw_estimands(design)
run_design(design)
summary(design)
compare_designs(design_1, design_2)
```

Design Diagnosis

Diagnose the properties of your design

```
diagnosis <- diagnose_design(
  design, sims = 100, bootstrap_sims = 100)
```

```
summary(diagnosis)
get_diagnosands(diagnosis)
get_simulations(diagnosis)
```

Custom diagnosands

```
diagnose_design(
  design,
  diagnosands = declare_diagnosands(
    sig_pos = mean(p.value < .05 & estimate > 0)))
```

CC BY SA Graeme Blair, Jasper Cooper, Alexander Coppock, Macartan Humphreys and Neal Fultz • declaredesign.org • Learn more at declaredesign.org/r/DeclareDesign • package version 0.18.0 • Updated: 2019-04

Phase A:

What is in the framework?

- An ontology: What exist
- A semiology: How signs / token represent things
- A grammar (Semantics): How to make meaningful expressions (using the signs / tokens)
- An epistemology: How knowledge works

These are explained in the slides that follow

Ontology: What is

- States
  - End
  - Intermediate
- Things
  - Compounds > Materials > Products > Product systems
  - Organisms > Landscapes > Biomes > Systems
- Agents
  - Individuals, Organisations or Movements
- Activities or processes
  - Decisions, actions, omissions
- Locality (spatio-temporal)
- Actuality and Potentiality (what is vs. what could have been)

Semiology: How signs refer to things

'Sense' and 'Reference'

Everything is a token

Quality of representation

- Validity
- Reliability
- Accuracy
- Repeatability

Semantics: Meaningful expressions

Activity: Agent - Intention - Thing - State

Event: Thing - State

Impact: State due to activity compared to likeliest alternative course of action

Epistemology: What you know

Observations

Instruments measurements reflect an aspect of the phenomenon or parameter of interest

Observations yield data

Constructs

Transformations and Statistics

Data is transformed into statistics and estimates through repeatable procudures

Causal models: How things are assumed to interact

Parameters to state Degrees of causality (scope 1-3)

Causal models give rise to methods for making claims about

Quality of knowledge

Same process at the level of data, estimates and models:

- Validity
- Reliability
- Accuracy
- Construct validity

Usefull knowledge is

- Relevant
- Complete
- Consistent
- Transparent
- Accurate