RISKCEDE

Services


process
projects
based
framework
programme
notifications
influential
participant
admin
riskcede
results
solution
time
polygon
specific
text
set
reporting
design
documents
targeted
database
input
poisonous
feedback
platform
reports
map
easy
mining
device
tree
report
models
plot
series
real
topic
e.g
details
cont
phone
system
outcomes
application
automated
notification
app
modelling
location
date
step
words
user
event
decision
flow
mobile
project
makes
issue
applied
users
knowledge
inputs
parameters
gathering
analysis
captured
information
data

Introduction

RiskCede created a platform to serve as an end-to-end solution for the data science pipeline, beginning with data gathering followed by tidying and wrangling of the data, then modelling and visualising before finally reporting.

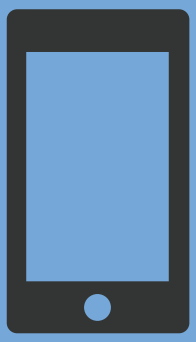
This framework includes an additional mobile app for capturing and validation of data. Using a smart phone as the source, login details, time and location are automatically detected from the device. It makes it easy to gather structured data, images and text.

The services project introduces RiskCede's solution for easy and accurate data gathering, while still utilising a powerful analysis and reporting platform.

The framework allows for fast implementation of new projects across any domain and works on all platforms.

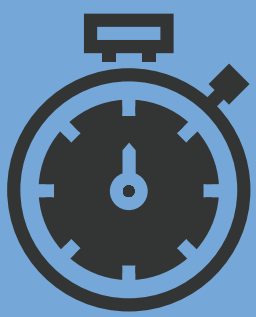
The aim of this project is to extract knowledge from the data, if possible, and use this knowledge to improve processes and procedures, in other words to 'operationalise' the mined knowledge. To achieve this there are two focus areas, the input data and the mining techniques, both of these are described in detail in this document.

Concept



Use mobile device to capture data, validate at input.
iOS or Android.
Projects on server linked to user at authentication.
Available 24/7, real-time process flow.

Analyse data in real-time:
- User details.
- Date and time.
- Geolocation.
- Inputs, structured or free text.
- Images.



- Real-time analysis and feedback directly back to the mobile device.

- Notification on device.

- Online web interface for data-mining.

- User input for different scenario testing.

- Static reports based on user input.

- Can be automated to be emailed at specified times.

Analysis and Reporting

1 Reporting

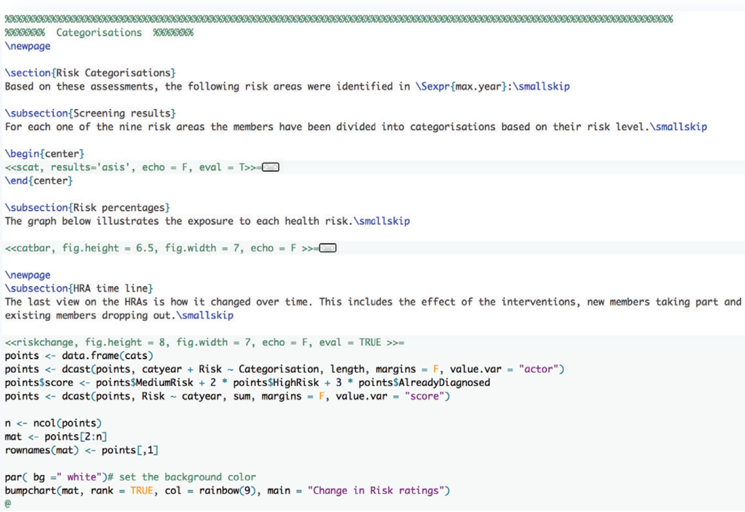
Reporting is based on reproducible research principals. This means the whole process of compiling the report is captured in one parameterised script. Thus everything from the data source, whether simply querying a database or scraping a website, right through to the document layout is specified in one script.

Rerunning the report is easy and quick and makes for fewer errors. It enables the creation of historic reports simply by changing the date input parameter.

Reports can also include dynamic portions, e.g. an additional paragraph to explain an anomaly if it occurs or an extra graph if it meets some specific criteria.

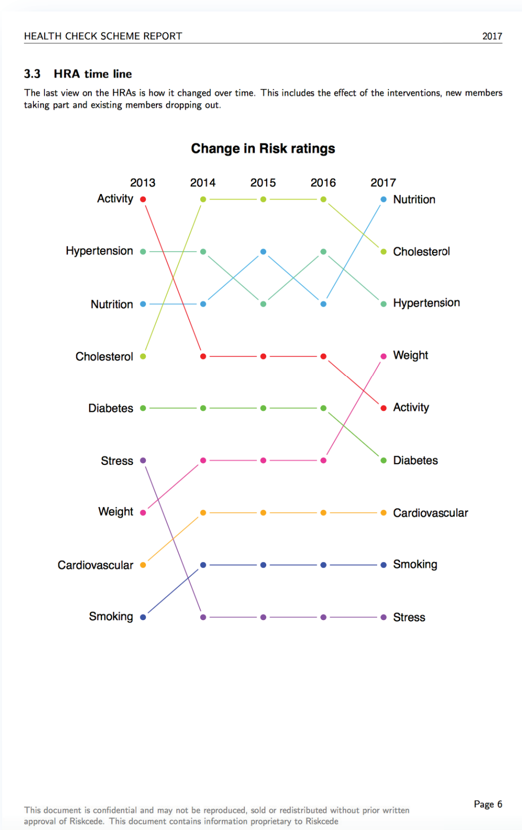
Reporting can be automated to run on a regular basis (e.g. an automated email each first day of the month) or when a specific event occurs, e.g. when a specific high or low value is captured. Reports can also be downloaded directly from a webpage, allowing the user to specify certain input parameters.

Input (Paramaterised)



```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%  Categorisations  %%%%%%%
\newpage

\section{Risk Categorisations}
Based on these assessments, the following risk areas were identified in \Sexpr{max.year}:\smallskip

\subsection{Screening results}
For each one of the nine risk areas the members have been divided into categorisations based on their risk level.\smallskip

\begin{center}
<<scat, results='asis', echo = F, eval = T>>=
\end{center}

\subsection{Risk percentages}
The graph below illustrates the exposure to each health risk.\smallskip

<<catbar, fig.height = 6.5, fig.width = 7, echo = F >>=

\newpage
\subsection{HRA time line}
The last view on the HRAs is how it changed over time. This includes the effect of the interventions, new members taking part and
existing members dropping out.\smallskip

<<riskchange, fig.height = 8, fig.width = 7, echo = F, eval = TRUE >>=
points <- data.frame(cats)
points <- dcast(points, catyear + Risk ~ Categorisation, length, margins = F, value.var = "actor")
points$score <- points$MediumRisk + 2 * points$HighRisk + 3 * points$AlreadyDiagnosed
points <- dcast(points, Risk ~ catyear, sum, margins = F, value.var = "score")

n <- ncol(points)
mat <- points[2:n]
rownames(mat) <- points[,1]

par( bg =" white")# set the background color
bumpchart(mat, rank = TRUE, col = rainbow(9), main = "Change in Risk ratings")
@
```

Result



HEALTH CHECK SCHEME REPORT 2017

3.3 HRA time line

The last view on the HRAs is how it changed over time. This includes the effect of the interventions, new members taking part and existing members dropping out.

This document is confidential and may not be reproduced, sold or redistributed without prior written approval of Riskcede. This document contains information proprietary to Riskcede

Page 6

Analysis and Reporting (cont.)

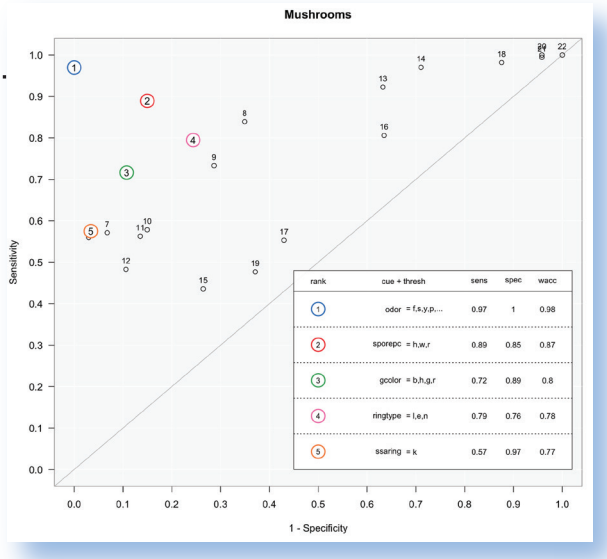
2 Machine learning

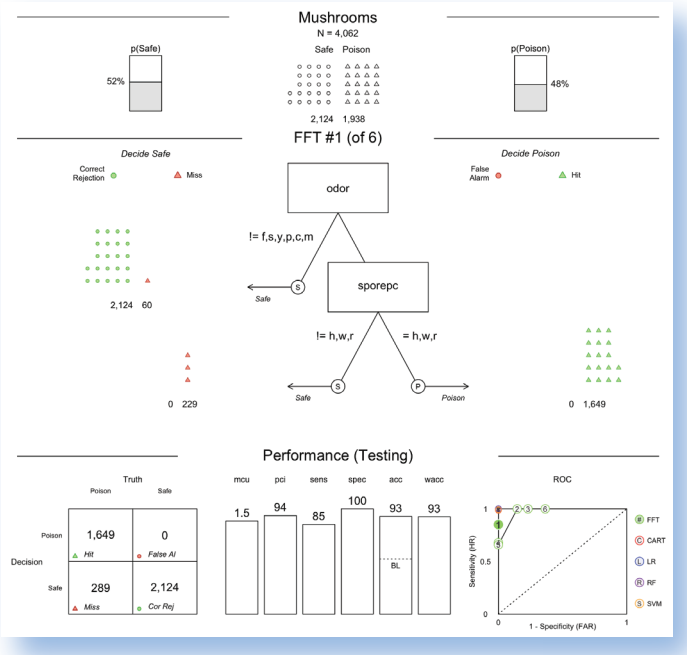
The following example illustrates the immediate feedback principle, using the FFTrees package decision tree model. In this project the user is assisted to determine whether a mushroom is edible or poisonous.

Based on research data containing 22 parameters of over 8000 mushrooms a decision tree algorithm is used to determine the likelihood of a mushroom being poisonous.

Firstly all the parameters are analysed. From the plot it is clear that the following five inputs are the most important.

1. odour: almond, anise, fishy, foul, etc.
2. spore-print-colour: black, brown, buff, etc.
3. gill-colour: black, brown, chocolate, etc.
4. ring-type: cobwebby, evanescent, flaring, etc.
5. stalk-surface-above-ring: fibrous, scaly, silky, smooth





A programme is then designed for the mobile app to capture this information and give back a result based on the user input.

The following decision tree is used to analyse the input in real-time and provide immediate feedback to the mobile device on whether the mushroom is poisonous or edible.

Analysis and Reporting (cont.)

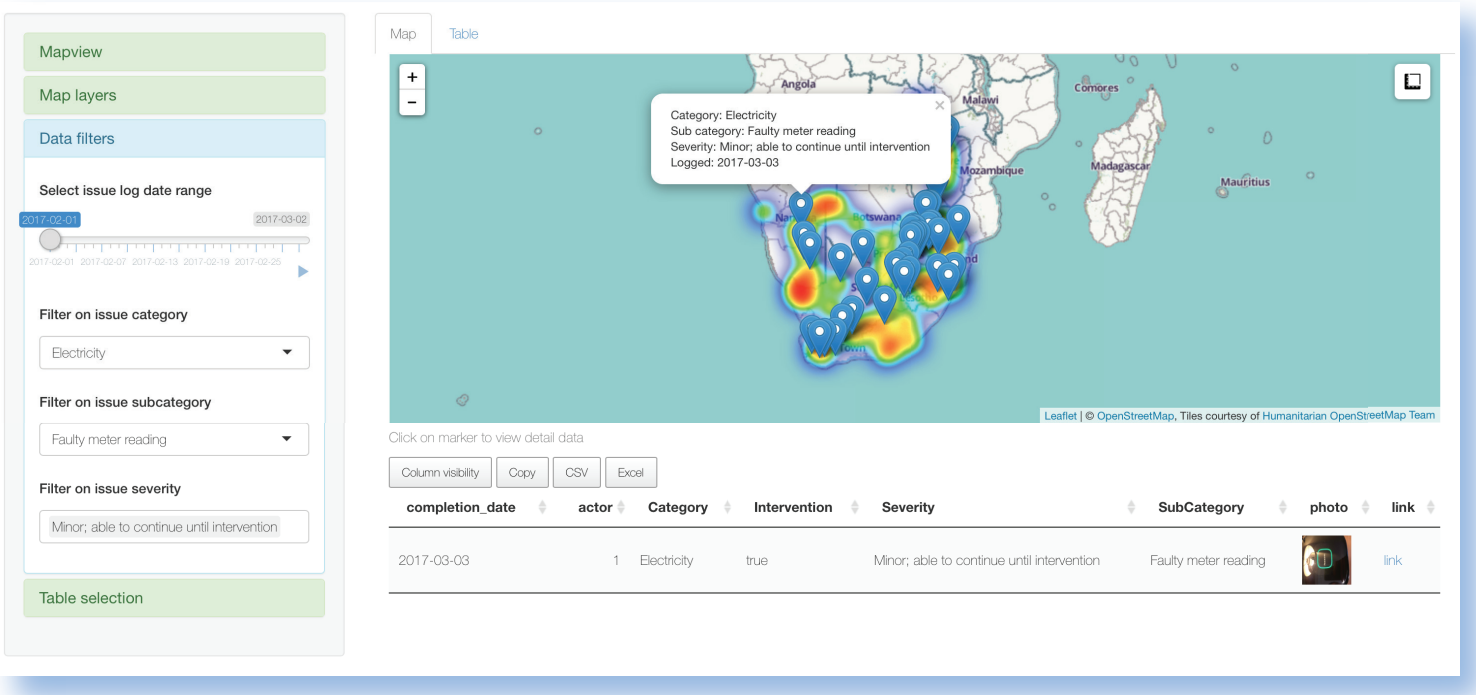
3 Geographic location

Data captured on the mobile device carries the GPS coordinates of the user, thus making geo-spatial analysis possible.These maps can be viewed online, for interactive analysis, or in a static report based on some user input parameters.

Point in polygon analysis:
The first step is to define a polygon as an area with a boundary (a province, a town, a farm, a mine). The next step is to link specific information to that polygon (for a municipality it could be the contact details of the person responsible for road maintenance, or the process to follow to buy property).

Then when a participant uses the app and submits information, the location of the phone is mapped and if found to be in a specific polygon, the linked rules or data are applied and can be sent back to the phone or used for other analyses.

Other functions of this service are plotting points on a map, overlaying a heat map on a map, finding nearest locations or points, and tracking of routes.



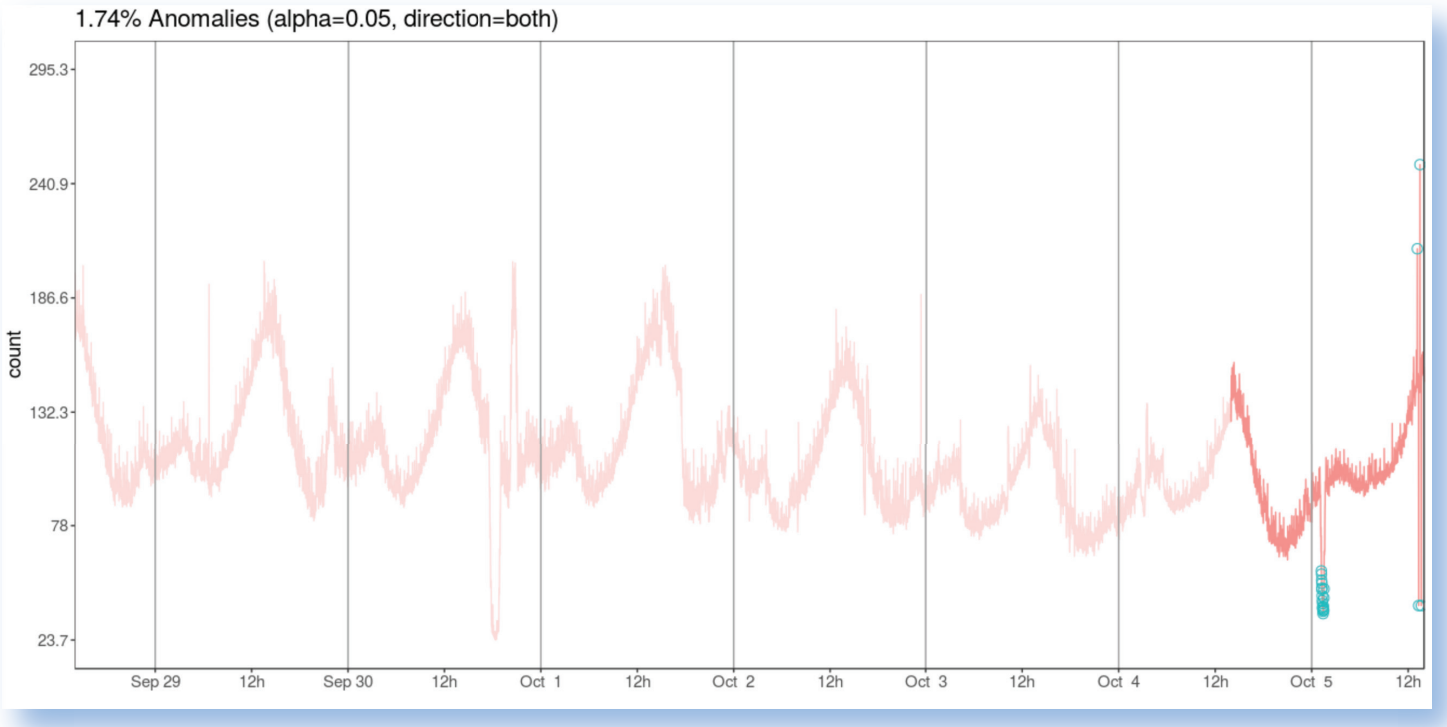
Analysis and Reporting (cont.)

4 Time series analysis

All data captured is time stamped with the date and time it was captured. This makes it possible to perform time series analysis on single value inputs or on results based on combined input, e.g. assigning an overall risk score based on all the captured inputs.

One area of interest might be the correlation between input and time, e.g. daily temperatures or time criminal activities are most likely to occur. Time series data is also used for predictive maintenance, to predict when an event will take place and then prevent it from happening.

These models identify events such as failures and then investigate the patterns in the time preceding that event. Models can be used to identify anomalies in the time series data for further investigation. These models are often applied to real time notifications to inform the user to perform an action.



Analysis and Reporting (cont.)

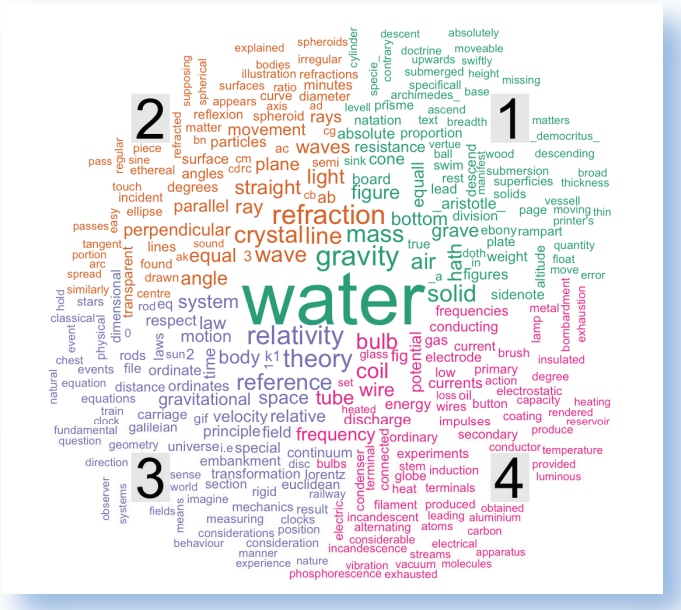
5 Text mining

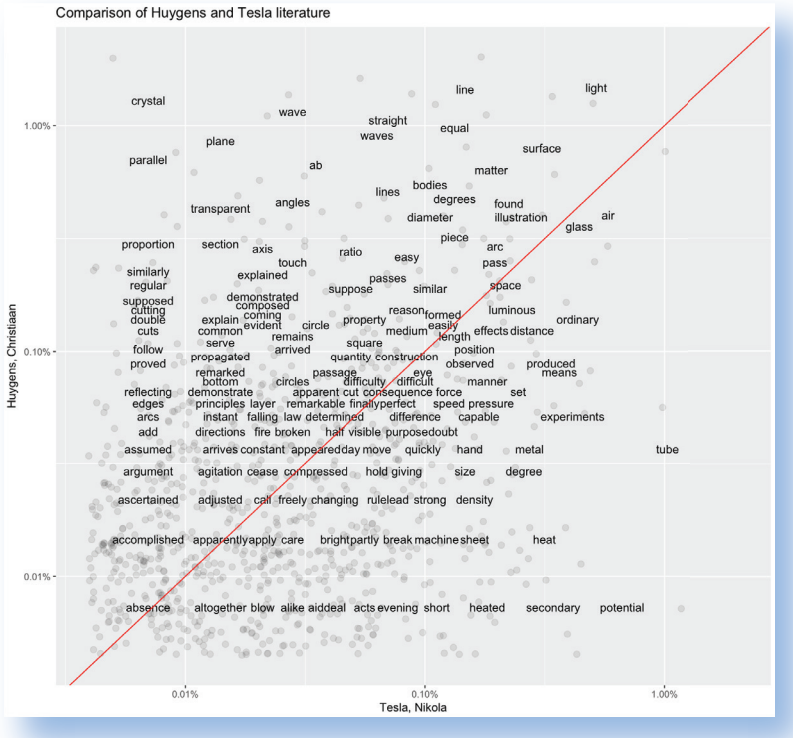
The app can also be used to gather free text (user opinions) about a topic. This in turn can be anaysed for key information, e.g. most influential words, sentiments, correlated words and topic modelling as well as comparing documents with each other.

Topic modelling creates natural clusters of words that occur together, thus a set of text will have a mixture of topics and each topic will be a mixture of words. It is useful to study these topics to determine the message of the text. Together with sentiment analysis it provides insight into the users' opinions about the issue addressed.

Sentiment analysis can also be used as a parameter, making it possible to view the inputs of participants with a positive sentiment versus those with a negative perception.

On the right is a list of words per topic based on a machine learning algorithm to find natural clusters in the text.





It is also possible to compare two sets of text with each other, to the left is a plot comparing the most influential words in two different documents. This plot illustrates the most influential words used in both documents, but also their relative frequency compared to each other.

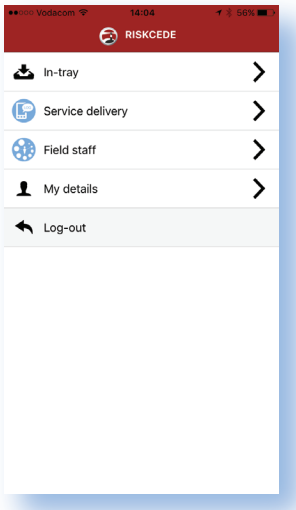
Mobile Application

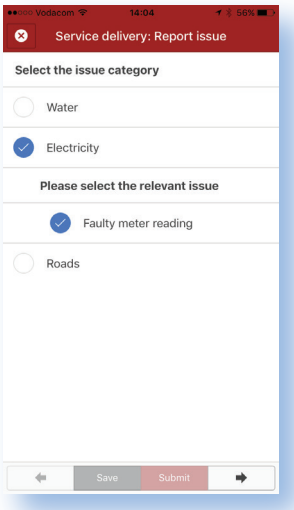
The mobile application builds up dynamically from the server workflow projects. This means that once a participant logs in, the application collects all projects assigned to that participant and displays it in his/her in-tray.

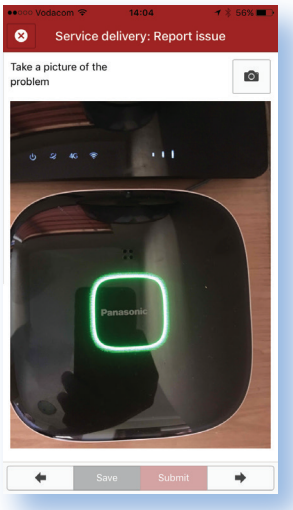
When a new project is launched, a notification is sent to the participant's phone (notifications are set up during the process design and can be set to be sent at any stage of the project).

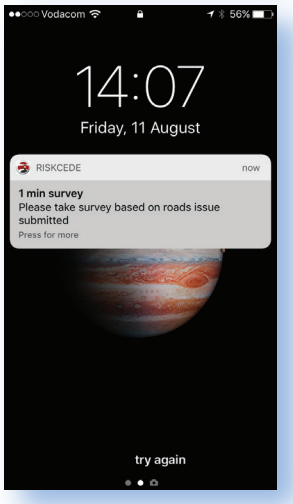
Some projects cycle continuously and are available again after completion, while others are removed from the participant's in-tray once they are completed.

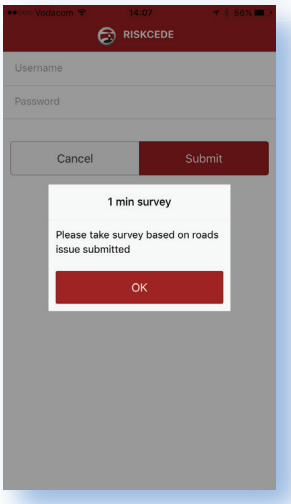
All data captured from the app contain information on the location and the time of the event as well as the user.









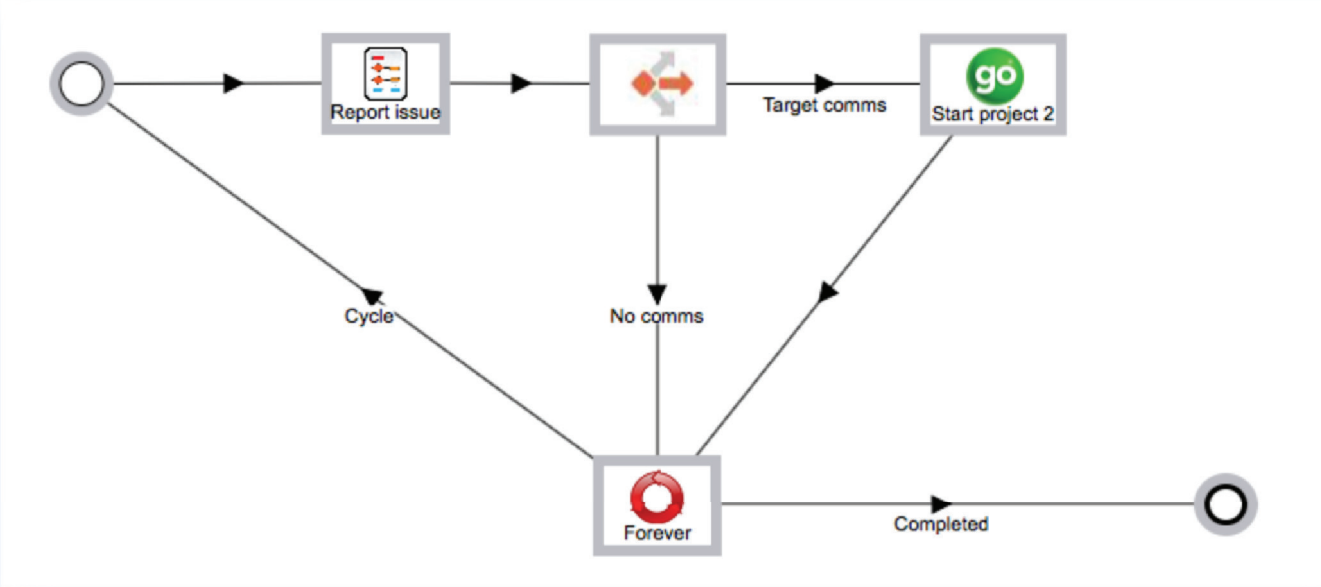


Process Flow

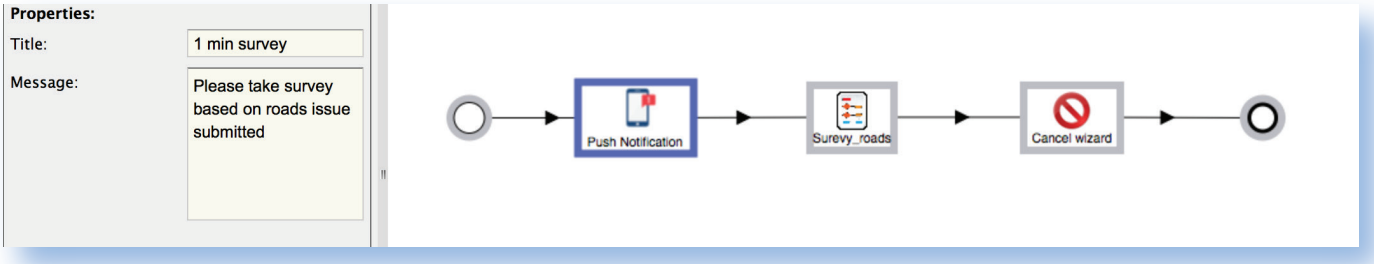
Once an instance is initiated for a participant a specific process flow follows, this includes events and their outcomes, decisions based on these outcomes, targeted communications and anything else the project requires to reach its goal. There are currently 44 of these step types available and new ones can be developed as necessary.

Below is the process flow for the demo service project:

This project starts with a programme that contains a questionnaire about the issue to be reported. Thereafter a targeted communication programme is launched for cases where the issue involved roadworks. This programme cycles continuously and is always available to the user.



The mobile app allows for notifications, a feature that is especially useful in automated projects. The project specifies the criteria for, and content of the notification and is sent for each instance at the correct time. This can serve as automated reminders, introduction of new projects or targeted communication.



Framework

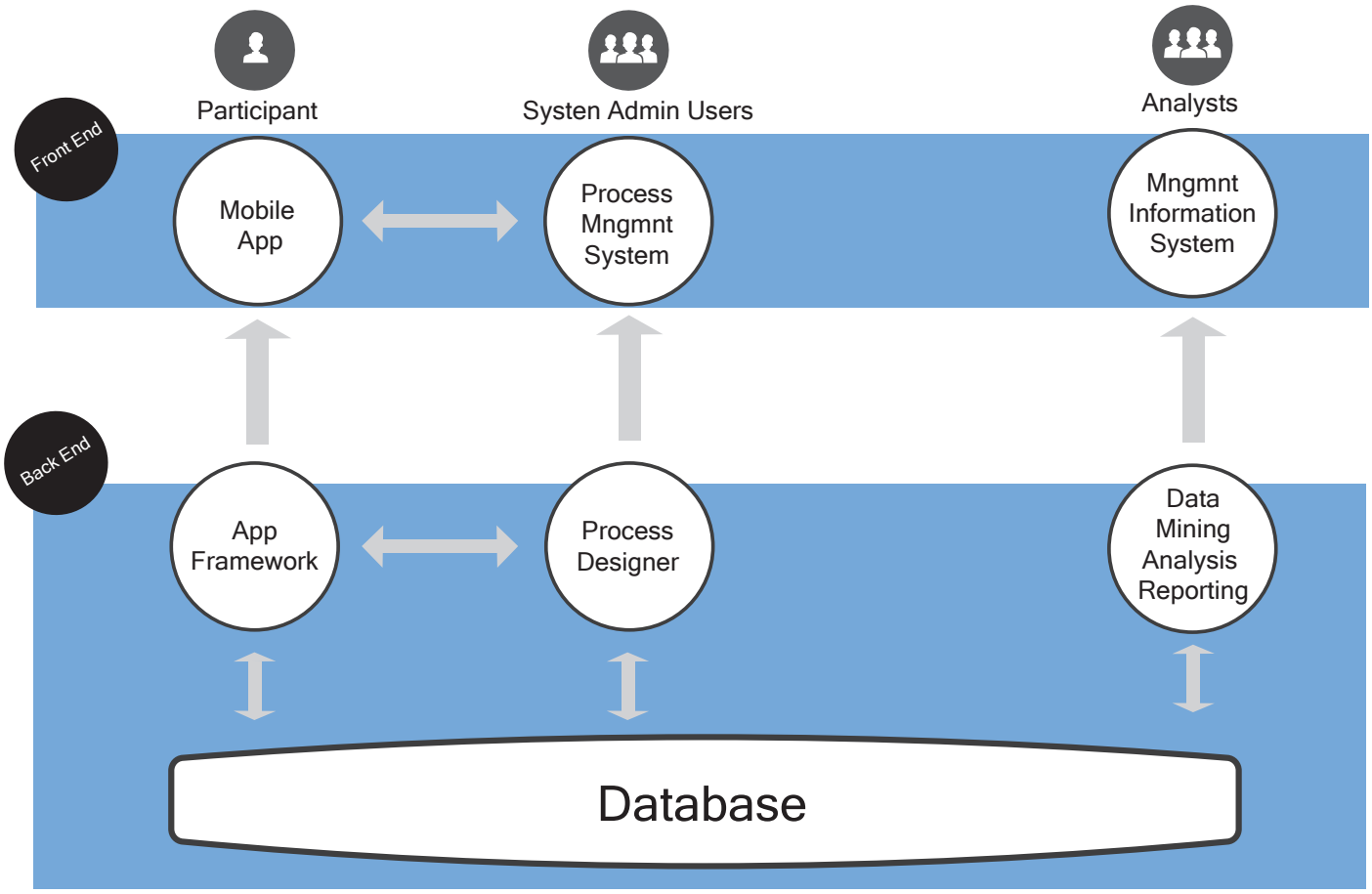
The software framework is designed from the database upwards, thus all front end applications (mobile and web admin pages) are rendered dynamically based on the database.

During design time the process flow designer is used to create projects with specific purposes/processes. At runtime the different instances follow these process patterns.

The reporting software uses this live data to report on, and to trigger warnings in predetermined outcomes.

Thanks to this design, it is easy to create and integrate new projects to the existing front end projects.

The final product has three users: the participant (mostly on the mobile app), the system admin users and the analysts.



Conclusion

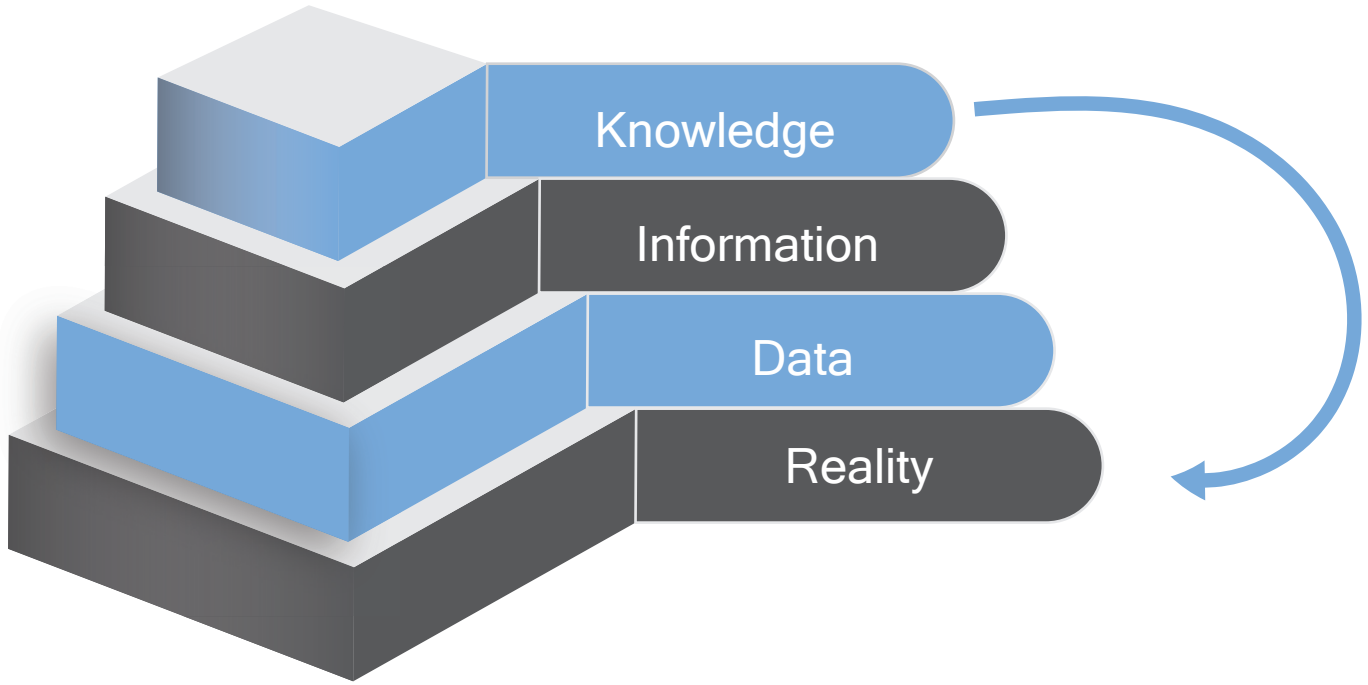
RiskCede uses analytics to make sense of real world data and uses a mobile app to make gathering the data and communicating the results easier and faster.

This single platform provides an end-to-end solution, from gathering information (and validating it at the point of entry to ensure quality of data) to performing predictive analysis and providing feedback, either in real-time or in report form for strategic decision-making.

The main advantage of this is access to instant (relevant) and accurate information. There is a lot less chance for error and reproducing an analysis or report is easy and effortless.

This platform is not domain specific and can be applied in any field and on any device.

The aim and purpose is to optimise decision-making through the best possible use of all available resources.



Results are applied to real life scenarios, either through strategies based on analysis results or real-time feedback.

Contact Us

info@riskcede.co.za
www.riskcede.co.za

RISKCEDE