

Decoding the Ancients

better practices and processes using R

Introduction

- **Logan Ashmore** – Senior Data Analyst,
Data Collection and Management
(Data & Evidence Secondment)

Presentation overview

- Improvements MfE is making in data collection and management
- Challenges to implementing change
- Contaminated Site Remediation Fund (CSRF) Example

Desired improvements

- **Data management and sharing**
- **Analysis, research, and reporting**
- **Reproducibility and
transparency**

Issues

- “Data” isn’t always data (if any exists)
- Building the plane as we fly
- Short attention spans
- Inherited products

Who are the Ancients?

- anyone no longer working at the organisation
- but we still use their product

What are we doing about it?

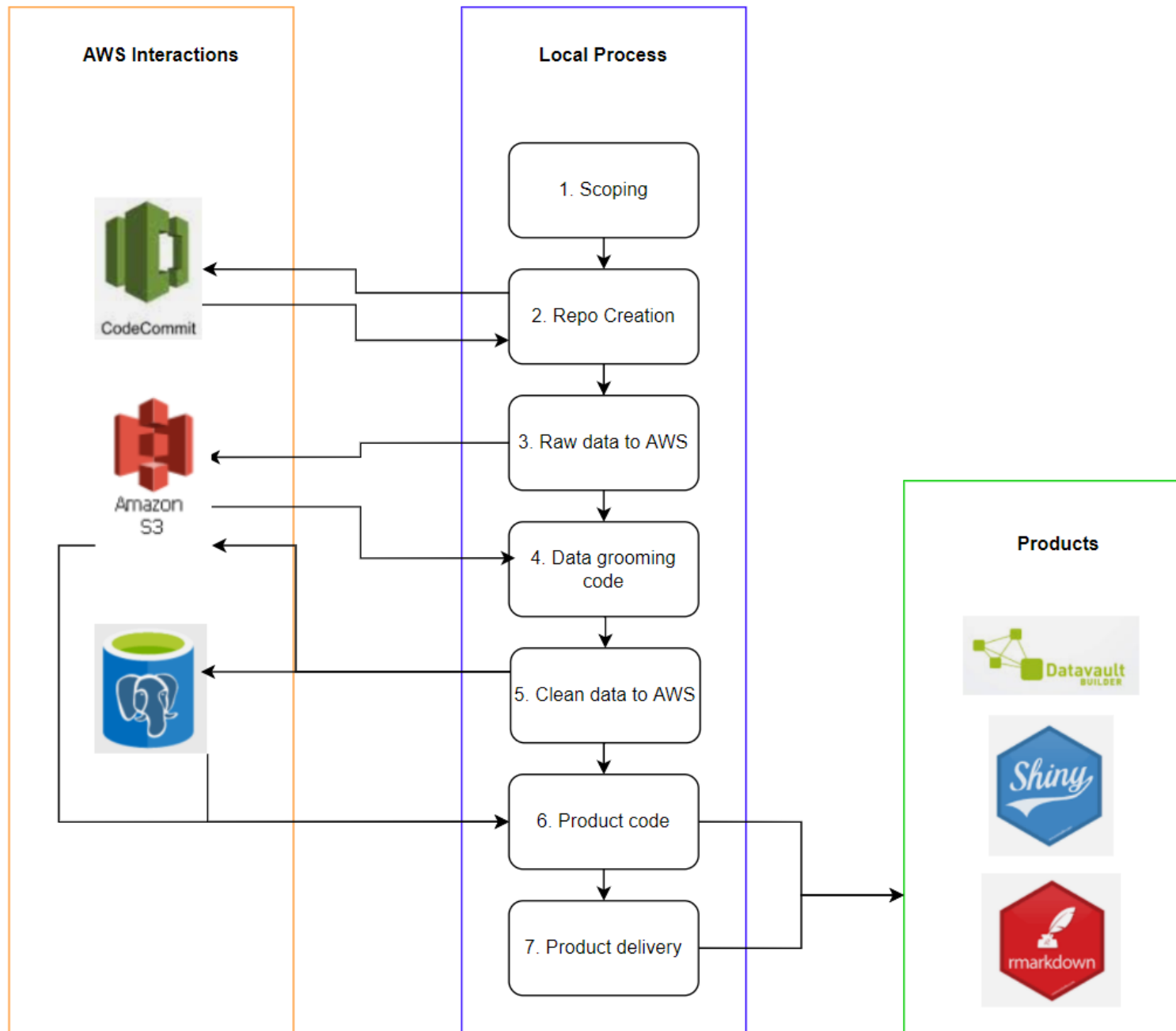
- **Upskilling in R (MfE)**
- **Converting old information sources to proper datasets**
- **Migrating to the cloud**
- **Building apps for teams to interact with the data**
- **Committing to Git**

What are we doing about it?

Still growing but have some firm ground rules:

- Always use Git
- Codecommit for internal projects, GitHub for external partnering/sharing
- Shiny Server for internal Apps, IO for public
- Repos are for code, S3 and RDS are for data
- No personal credential files in code or repo

Normal project progression



The CSRF Program

- 10 contaminated sites receive government funding every year for remediation
- Sites prioritised by highest risk
- MfE content
- Envirolink content

CSRF Tool

- **Procured project, 2011**
 - **Last meaningful update 2013**
- **Macro heavy excel 2007 workbook**
 - **Uses Monte Carlo simulations to score and rank risks**
- **Data entry, storage, reporting, and simulation all in one**
 - **more than one version**

CSRF Challenges

The image shows a screenshot of a complex spreadsheet, possibly a project management or financial tool. The spreadsheet is organized into several columns and rows, with various data points and formulas. The columns are labeled with various categories, and the rows contain numerical data. The spreadsheet is divided into several sections, with some columns highlighted in yellow and others in blue. The data appears to be organized into a grid, with some cells containing text and others containing numbers. The overall layout is dense and detailed, suggesting a high level of complexity in the data being presented.

CSRF Challenges

Contaminated Site Remediation Fund Prioritisation Tool															
CSRF23 Kaipapepe Canal / Whakaitane															
Calculate Site Score															
Spell Check															
Human Health						Ecological receptors									
	Min	Mean	Max	Comments		Min	Mean	Max	Comments						
Soil	Toxicity	AP	10	AP	Dissolve in sediments Length of the canal Direct contact with sediment in canal swimming. Possible direct exposure via fish (or shell). Small size of contaminated shell in canal bank involves contact with larger amount in canal. occasional contact, primarily fishing, but swimming estimated at 50 days/year. Small children unlikely to be exposed. School children and adults.	Dissolve in sediments	Toxicity	AP	10	AP	Terrestrial (Plants or Animals)	Toxicity	AP	10	AP
	Amount	AP	10	AP			Amount	AP	10	AP		Amount	AP	10	AP
	Exposure	AP	10	AP			Exposure	0	0	0		Exposure	0	0	0
	Duration	2	2	5			Duration	--	Duration	--		--	Duration	--	--
Receptor	3	6	6	Receptor	9	9	9	Receptor	--	--	Receptor	9	9		
Water Use	Toxicity	AP	10	AP	Dissolve Large length of canal effects Unknown, but is not valued or a concern in any reports. Dissolve leachability, highly immobile in ground water.	Dissolve	Toxicity	AP	10	AP	Aquatic - Overland pathway	Toxicity	AP	10	AP
	Amount	AP	10	AP			Amount	AP	10	AP		Amount	AP	10	AP
	Exposure	0	0	0			Exposure	AP	10	AP		Exposure	AP	10	AP
	Duration	--	Duration	--			Duration	5	10	AP		Duration	5	10	AP
Receptor	--	Receptor	6	6	Receptor	1	3	6	Receptor	24	24	24			
Homegrown Vegetable	Toxicity	--	Toxicity	--	On-site Receptors	On-site Receptors Only	Toxicity	AP	10	AP	Aquatic - Groundwater	Toxicity	AP	10	AP
	Amount	--	Amount	--			Amount	AP	10	AP		Amount	AP	10	AP
	Exposure	0	0	0			Exposure	--	Exposure	--		--	Exposure	--	--
	Duration	--	Duration	--			Duration	--	Duration	--		--	Duration	--	--
Receptor	--	Receptor	6	6	Receptor	--	Receptor	6	6	Receptor	6	6			
Other Produce Consumption (e.g. Livestock, Eggs, Fish)	Toxicity	AP	10	AP	Dissolve Large length of canal effects, but has a ditch or great or SKM calculator - see comment below. Dissolve accumulated in soil. Depends on where you are caught from. Likely that exposure is an average, not max. Eggs known to be taken for food within the Hauri community. Eggs eaten frequently in the community by young people to avoid from community ditches if taken from canal. Unknown whether eaten by young children (toddlers). Certainly by older children and adults. Note, SKM is not a risk - use more conservative USEPA slope factor rather than Hauri approved interim maximum monthly intake of 30 µg/kg bw/d. Results in remediation grade being less than by a factor of around two, or more depending on whether take their max or average scenario.	Dissolve	Toxicity	AP	10	AP	Cultural Social Heritage Economic benefit	Toxicity	AP	10	AP
	Amount	5	8	0			Amount	5	8	0		Amount	5	8	0
	Exposure	AP	10	AP			Exposure	AP	10	AP		Exposure	AP	10	AP
	Duration	0	0	AP			Duration	0	0	AP		Duration	0	0	AP
Receptor	3	6	AP	Receptor	3	6	AP	Receptor	3	6	AP				
Score						Coefficient of Variation									
Human Health						Human Health + Ecological									
Ecological						Cultural									
Human Health + Ecological						Social									
Cultural						Heritage									
Social						Economic benefit									
Heritage						Economic benefit									
Economic benefit						Economic benefit									

CSRF Model

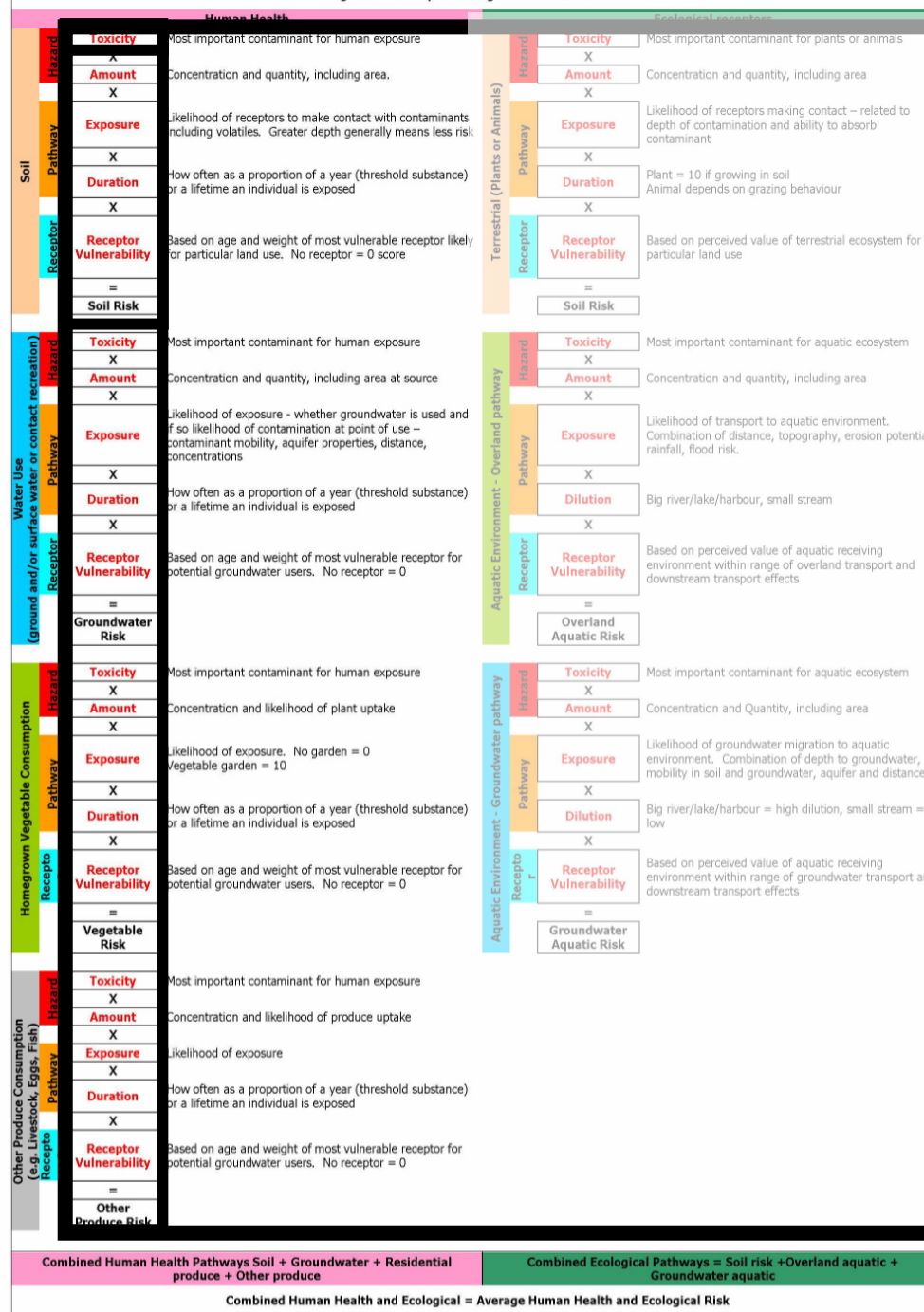
Modal risk (1-10)

Product of risks /1000 (Exposure)

Sum of Exposures (indicator score)

Average of Indicators (overall score)

Figure 1 – Conceptual Design of Prioritisation Tool



Shiny App

- **Replicated style and language of the spreadsheets**
 - **ease transition for users**
- **Leveraged Apply family and custom functions to generate repetitive UI elements and reactivities**
- **Created additional data set to store historic edits**

Shiny App

CSRF Site Prioritisation Tool

Select the site to edit:
CSRF23 - Kopepeo Canal Upload Edits Store Edits Remove Entry Site Priority- Ranking **4** Score **37**

Site Information **Parameter Entry** Rankings Tracking Info Tables

Update Scores

Social-Cultural Impacts

Score: **0**

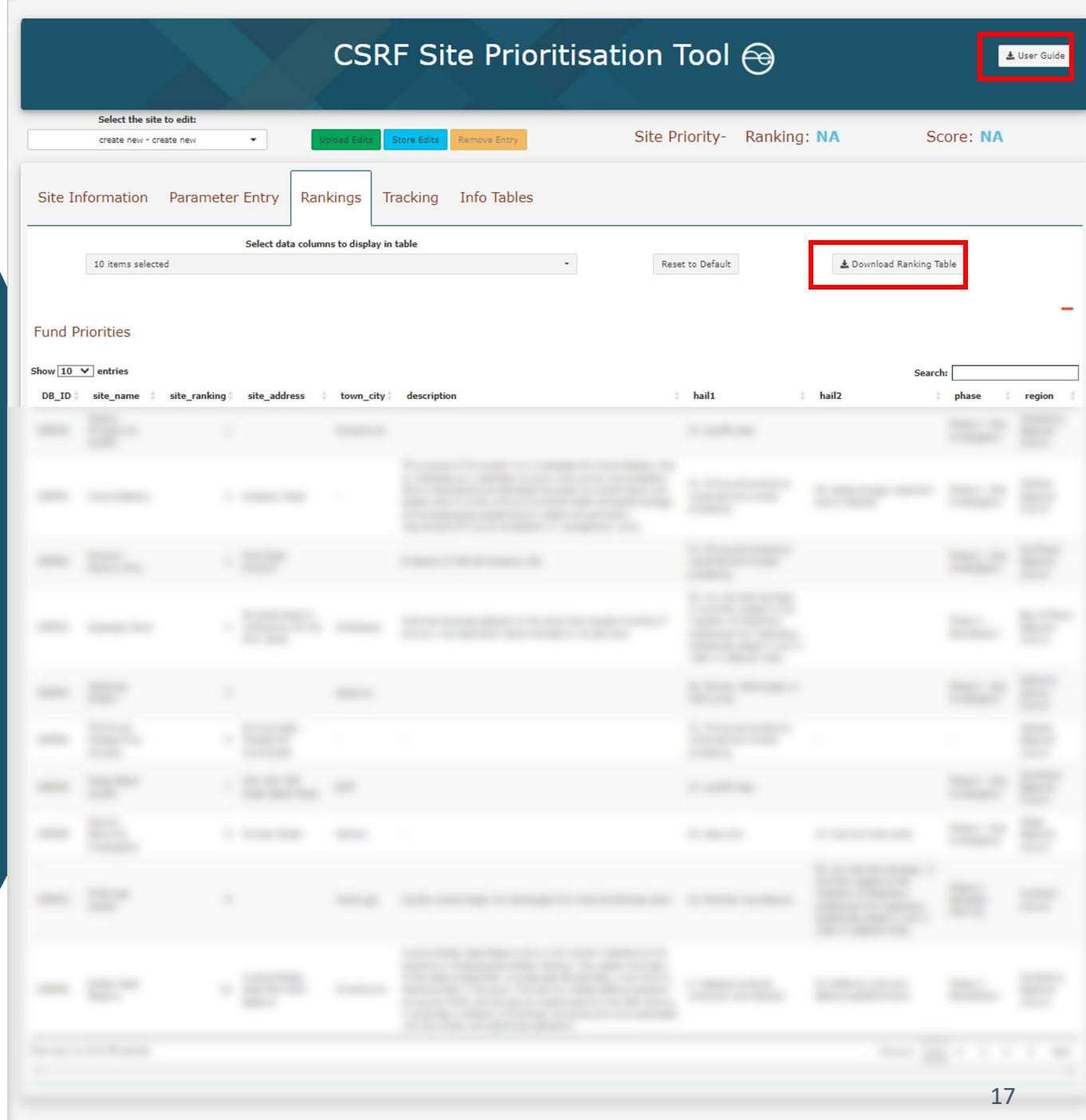
Cultural None	Social None	Heritage None	Economic Benefit None
Comment:	Comment:	Comment:	Comment:

Human Health

Score: **48**

			Min:	Mean:	Max:	Comment:
Soil Score: 12	Hazard	Toxicity	10	10	10	dioxins in sediments
		Amount	10	10	10	length of the canal
	Pathway	Exposure	10	10	10	Direct contact with sediment in canal if swimming.
		Duration	2	2	5	occasional contact, primarily fishing, but swimming
Receptor		3	6	6	Small children unlikely to be exposed. School children	
Water Use Score: 0 Onsite Water User: FALSE	Hazard	Toxicity	10	10	10	Dioxins
		Amount	10	10	10	Large length of canal affected
	Pathway	Exposure	0	0	0	unknown, but is not voiced as a concern in any reports.
		Duration	-	-	-	
Receptor		-	-	-		
Homegrown Vegetable Score: 0	Hazard	Toxicity	-	-	-	
		Amount	-	-	-	

Shiny App



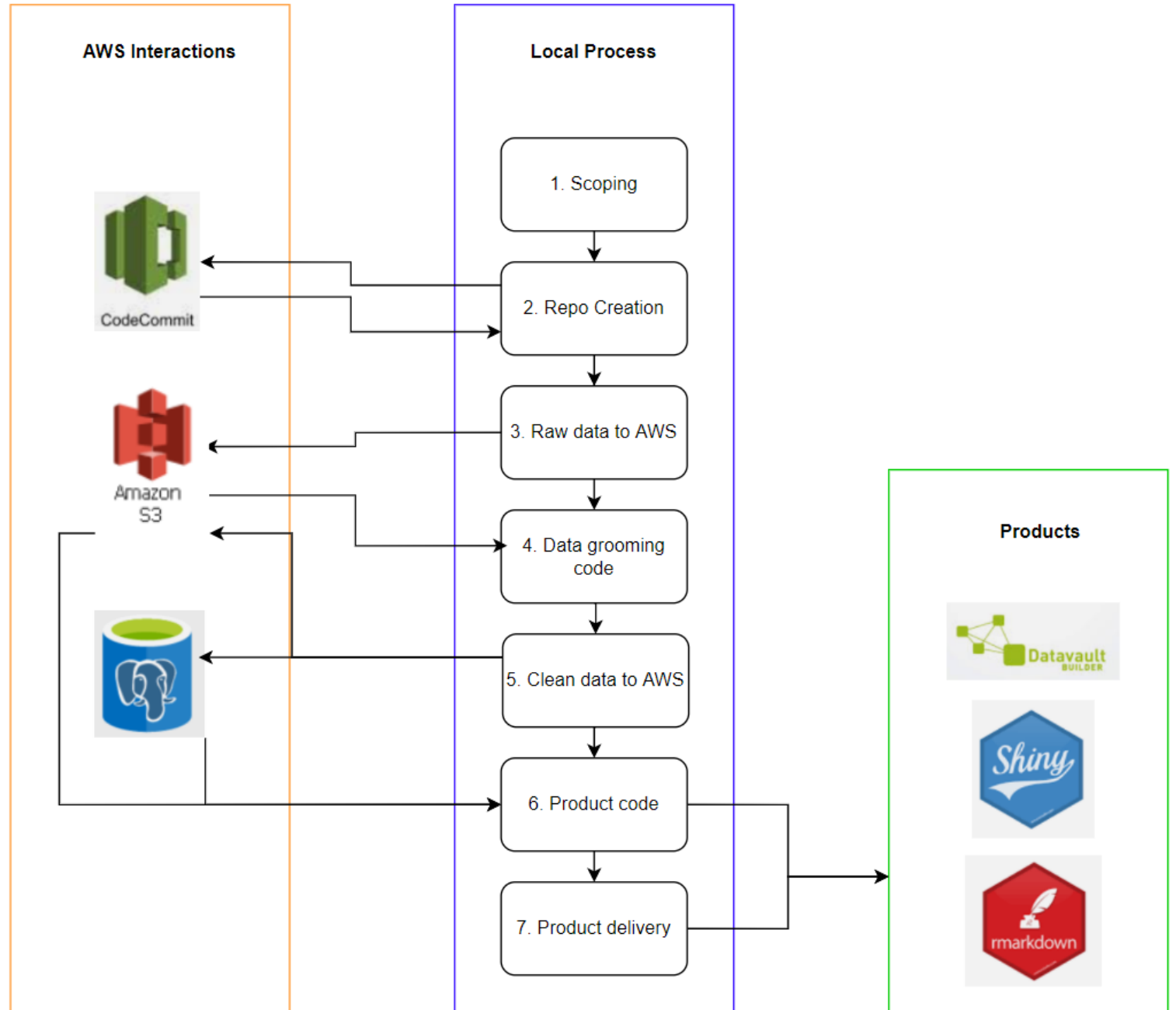
The screenshot displays the 'CSRF Site Prioritisation Tool' interface. At the top, there is a dark blue header with the tool's name and a logo. A 'User Guide' button is highlighted with a red box. Below the header, a navigation bar includes a dropdown menu for 'Select the site to edit' (currently showing 'create new - create new'), buttons for 'View Entry', 'Score Entry', and 'Remove Entry', and status indicators for 'Site Priority-', 'Ranking: NA', and 'Score: NA'. A secondary navigation bar contains tabs for 'Site Information', 'Parameter Entry', 'Rankings' (which is active), 'Tracking', and 'Info Tables'. Below this, a 'Select data columns to display in table' dropdown is set to '10 items selected', with a 'Reset to Default' button and a 'Download Ranking Table' button highlighted with a red box. The main content area is titled 'Fund Priorities' and features a 'Show 10 entries' dropdown and a search bar. A data table is displayed with the following columns: DB_ID, site_name, site_ranking, site_address, town_city, description, hail1, hail2, phase, and region. The table contains several rows of data, with the first row being partially visible.

DB_ID	site_name	site_ranking	site_address	town_city	description	hail1	hail2	phase	region
...
...
...
...
...
...
...
...
...
...

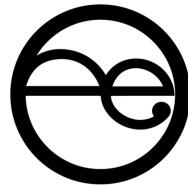
User Guide

- Produced using Rmarkdown
- Word output, reference MfE template
- Time cost upfront,
 - but easier to update in future
 - Cleaning/processing scripts, app, and guide all in one repo

User Guide



Questions?



Ministry for the
Environment
Manatū Mō Te Taiao