BIOWHERE – GEOREFERENCING NEW ZEALAND'S BIOTA FROM TEXTS
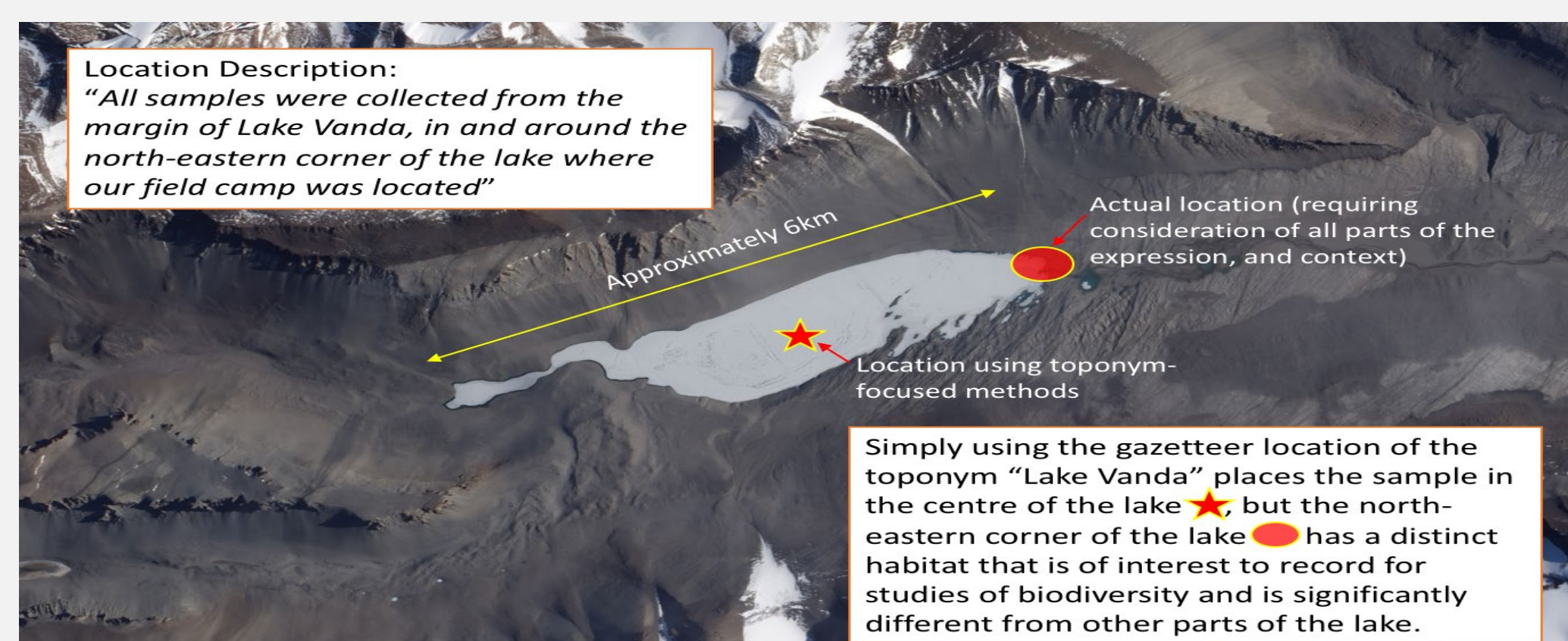
Kalana Wijegunarathna[1], Kristin Stock[1], Christopher B. Jones[2], Pragyan Das[1], Aron Wilton[3], Jonathan Procter[4], Hone Morris[5], David Medyckyj-Scott[3], Fraser Morgan[3], John Wieczorek[4] and Brandon Whitehead[3]

[1]Massey Geoinformatics Collaboratory – Massey University, New Zealand
[2]School of Computer Science and Informatics – Cardiff University, United Kingdom
[3]Manaaki Whenua Landcare Research, New Zealand
[4]School of Agriculture and Environment – Massey University, New Zealand
[5]Te Pūtahi-a-Toi – Massey University, New Zealand
[6]Rauthiflor LLC

Introduction

- Over 12 million records of biota specimens have been collected in New Zealand and Antarctica in the past 250 years.
- These records are georeferenced textually across various scientific publications and specimen collections held by museums and other institutions worldwide.
- BioWhere will develop cutting-edge methods to map this large volume of biota specimens.

Location Description:
"All samples were collected from the margin of Lake Vanda, in and around the north-eastern corner of the lake where our field camp was located"

Approximately 6km

Actual location (requiring consideration of all parts of the expression, and context)

Location using toponym-focused methods

Simply using the gazetteer location of the toponym "Lake Vanda" places the sample in the centre of the lake ★, but the north-eastern corner of the lake ● has a distinct habitat that is of interest to record for studies of biodiversity and is significantly different from other parts of the lake.
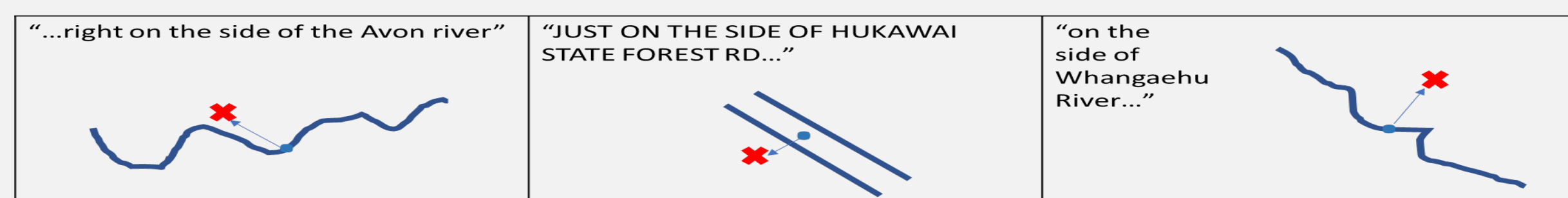
Self Learning Gazetteer

- We are designing and implementing a gazetteer that comprehensively records NZ's toponyms including Māori and Pākehā place names.
- The gazetteer will accommodate alternative, colloquial, multilingual and historical names, enabling the georeferencing of specimen records collected over nearly three centuries.
- Incorporating physical, historical, and cultural context will enrich the gazetteer with Māori knowledge including the origin, narrative and meaning of Māori place names.
- Equipped with our modern machine learning models, the self learning gazetteer will be able to map many current and historical place names accurately for the very first time.
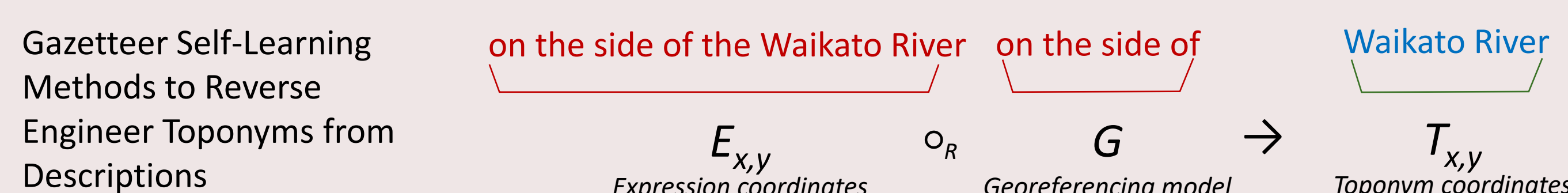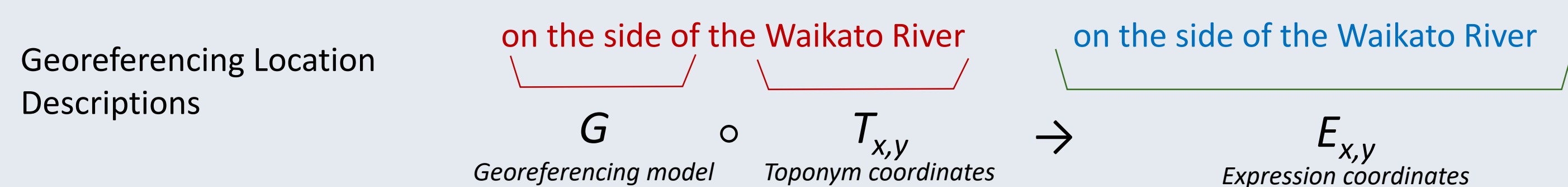
Georeferencing methods

- Going beyond simple toponym lookup.
  E.g. – "Ruatangata West, on roadside, 0.7 km along Worsnops Road from Ruatangata, east side of road.."
- Complex, multiclausal descriptions often contain vague spatial relation terms (here on, along, east) that can differ greatly in their interpretation according to the context.

"...right on the side of the Avon river"

"JUST ON THE SIDE OF HUKAWAI STATE FOREST RD…"

"on the side of Whangaehu River…."

- We will adapt the latest transformer-based machine learning architectures, to learn simultaneously from a large corpus of previously-annotated texts and from associated contextual data.
- Our dynamic, contextual model considers multiple factors to infer geographical acceptance areas including:
  - Environmental factors (e.g. habitat and climate);
  - linguistic context, including words that accompany spatial relations;
  - named place characteristics (e.g. geographic feature type, population);
  - situational context such as collector and date

Reverse Engineering toponym coordinates

Georeferencing Location Descriptions

on the side of the Waikato River

$$G \circ T_{x,y} \rightarrow E_{x,y}$$

Georeferencing model   *Toponym coordinates*   *Expression coordinates*

on the side of the Waikato River

Gazetteer Self-Learning Methods to Reverse Engineer Toponyms from Descriptions

on the side of the Waikato River   on the side of   Waikato River

$$E_{x,y} \circ_R G \rightarrow T_{x,y}$$

Expression coordinates (from collections databases)   *Georeferencing model*   *Toponym coordinates*

Reverse-Engineering Example

Impacts and benefits

- Understand species distributions through time and space – protect endangered species, pest control
- Multiple phenomena (e.g., health, disasters) can be efficiently mapped from social media posts, blog posts or archives – disaster management, environment management.
- Search tools across applications can better locate places using Māori and Pākehā place names – enabling Iwi and hapū to target environmental plans and policies.
- Increased opportunities for Māori in GIS and IT.

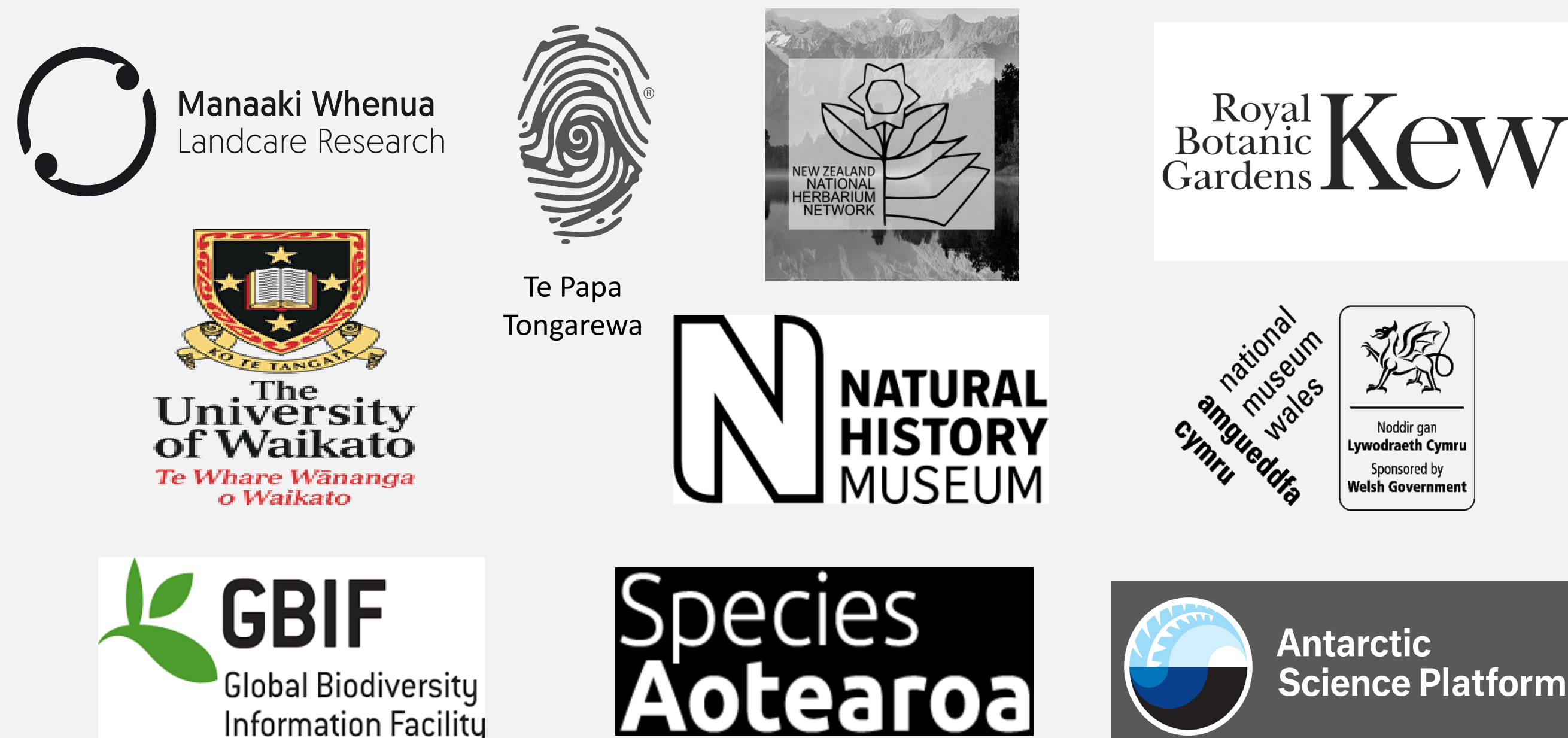Source: Manaaki Whenua-Landcare Research, Allan Herbarium

Source: Manaaki Whenua-Landcare Research, NZ Arthropod Collection

Source: Manaaki Whenua-Landcare Research, Allan Herbarium
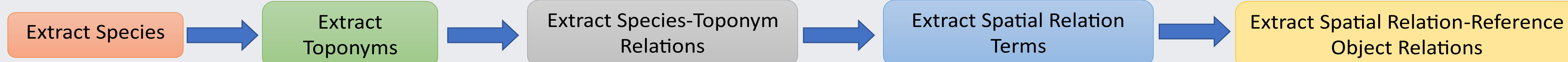
Our Partners

Manaaki Whenua Landcare Research
Te Papa Tongarewa
NEW ZEALAND NATIONAL HERBARIUM NETWORK
Royal Botanic Gardens Kew
The University of Waikato — Te Whare Wānanga o Waikato
NATURAL HISTORY MUSEUM
national museum wales — amgueddfa cymru — Noddir gan Lywodraeth Cymru Sponsored by Welsh Government
GBIF Global Biodiversity Information Facility
Species Aotearoa
Antarctic Science Platform

Extraction from collection database

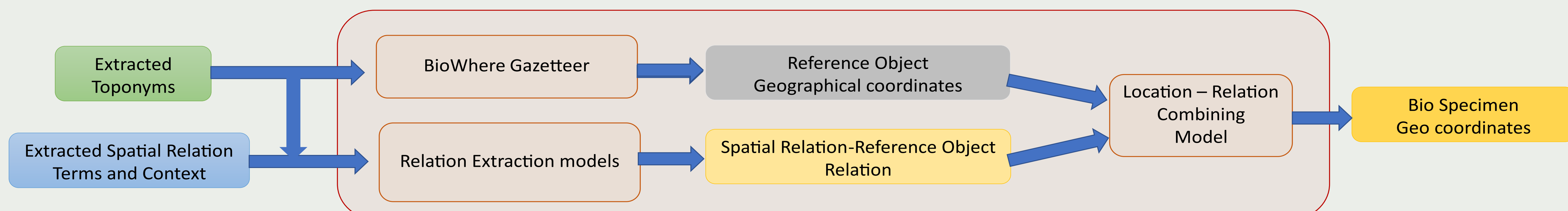| SpecimenGuid	LandDistrict	Locality	CollectionDate	SiVerbatimCollector
07C5969D-096C-4ED1-B81B-3C28121C4DBE	Southland Land District	on northern end of South Promontory, Snares Island	1971/02/18/	D.S. Horning. No 5A 66
9FDA9188-78E2-4198-A50E-576580961E19	Otago Land District	North Otago, eastern slopes of Dingle Burn, Lake Hawea	1990/02/26/	Lee WG
127C2997-83E0-4E6C-9D37-628BE13F0C3B	Hawke's Bay Land District	Hawkes Bay, Junction of Tutaekuri and Ngaruroro Rivers	1990/01/00/	Hartly J
084648A6-0D78-4245-ABF7-D8E8E31E566F	Canterbury Land District	Canterbury, Lower Rakaia River, N. bank near road and railway bridges	1970/03/08/	Parham BEV 12007
1BADE121-BA7E-4559-94E2-7E0BAFD918EC	Canterbury Land District	North Canterbury, Selwyn River near bridge on Springston-Leeston Rd	1970/02/19/	Parham BEV 11993
088E87E5-79C0-4111-B095-5C48E0C23E0A	Canterbury Land District	Canterbury, Banks Peninsula (near Okuti Valley Rd. and Usshers Rd)	1971/01/00/	P. Ussher
68B18E27-2B2A-414A-BC08-99FBAEE3E294	NULL	Whanganui, Wanganui City, corner of Dublin St and Victoria Ave	1989/10/16/	Ogle CC 1871
7F5790CB-2202-419A-BE72-7685F8D13218	NULL	North Island, Great Barrier Island, Old Clay track to Rosalie Bay	1989/11/22/	de Lange PJ 128
6CD04BF4-B1C2-410F-A76F-E6AAF06E1522	Southland Land District	Fiordland National Park: Murchison Mtns, summit of limestone ridge, N of Lake Orbell	1994/01/27/	Allan J. Fife 10329
634A3CB2-892C-431D-9868-EA257EA7FD23	Wellington Land District	Wanganui, east side of Great North Road, 35 to 55 m south of Parsons Rd intersection	2005/06/16/	Graeme La Cock
6392F80B-80B8-480A-A717-1E5D4198AF25	Southland Land District	Fiordland on a ridge between Lake Ione and the Phillipson Burn	NULL	Don Bruce

Example

on northern end of | South Promontory | Snares Island

North Canterbury | Selwyn River | near | bridge | or | Springston-Leeston Rd

| Method | Not required: species already stored as attribute in collections databases | NER + environmental variables | Not required: stored in database structure | Transfer learning: NER + habitat, landscape descriptions | Transfer learning applying relation extraction methods with environmental information |

Extract Species → Extract Toponyms → Extract Species-Toponym Relations → Extract Spatial Relation Terms → Extract Spatial Relation-Reference Object Relations

Georeferencing of Extracted specimens

Extracted Toponyms → BioWhere Gazetteer → Reference Object Geographical coordinates

Extracted Spatial Relation Terms and Context → Relation Extraction models → Spatial Relation-Reference Object Relation

→ Location – Relation Combining Model → Bio Specimen Geo coordinates

Self Learning Gazetteer