GOING BEYOND GOOGLE

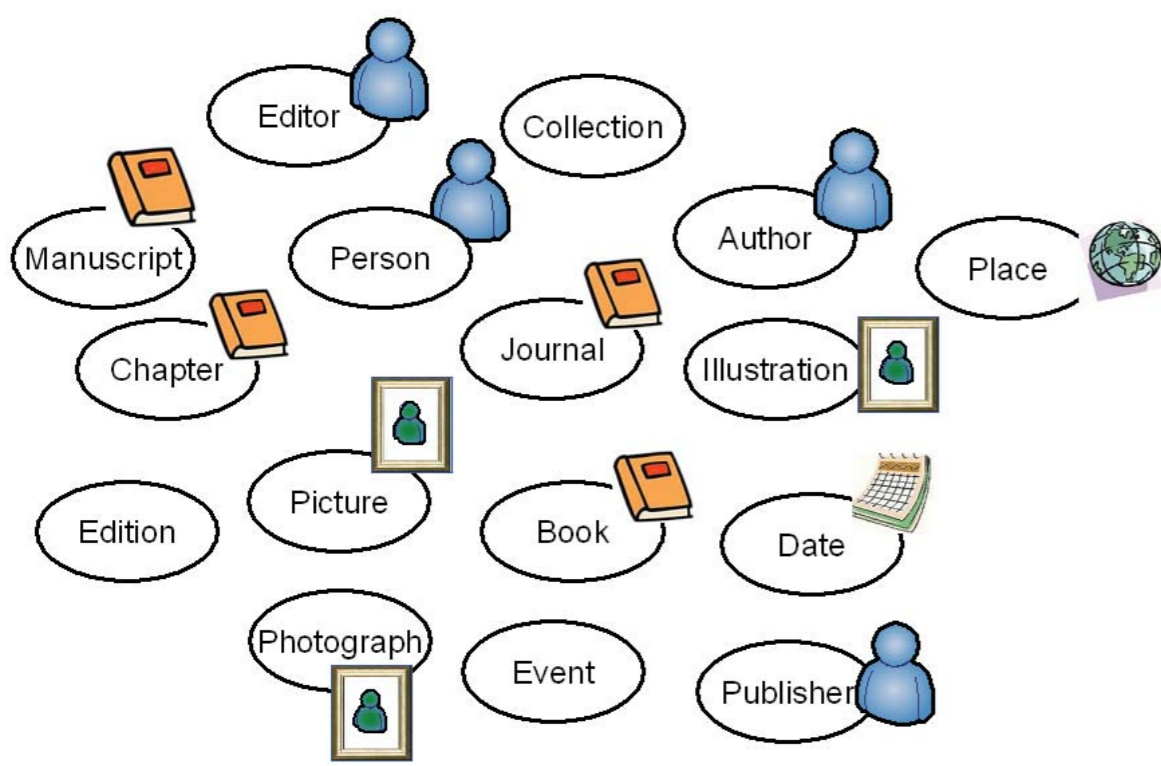## REPRESENTATION AND RETRIEVAL OF INFORMATION USING TOPIC MAPS

Library collections include increasingly large amounts of digital material. Libraries are digitising parts of their collection to provide wider access to important resources, and born-digital material is being added to collections. Basic text string searching and linear chapter-by-chapter browsing functionality is usually provided but so much more is possible.

The implicit linkages and cross-references between books, which occur in all print collections, can be made explicit in a collection of electronic texts. Correctly encoded, they create a framework to provide users with the ability to move horizontally between books and collections by following links between topics. Using this approach the New Zealand Electronic Text Centre (NZETC) has explored and developed an improved means of navigation for its growing online digital library using Topic Map technology. Like a simple back-of-book index or a library classification system, a topic map aggregates information to provide binding points from which everything that is known about a given subject can be reached.
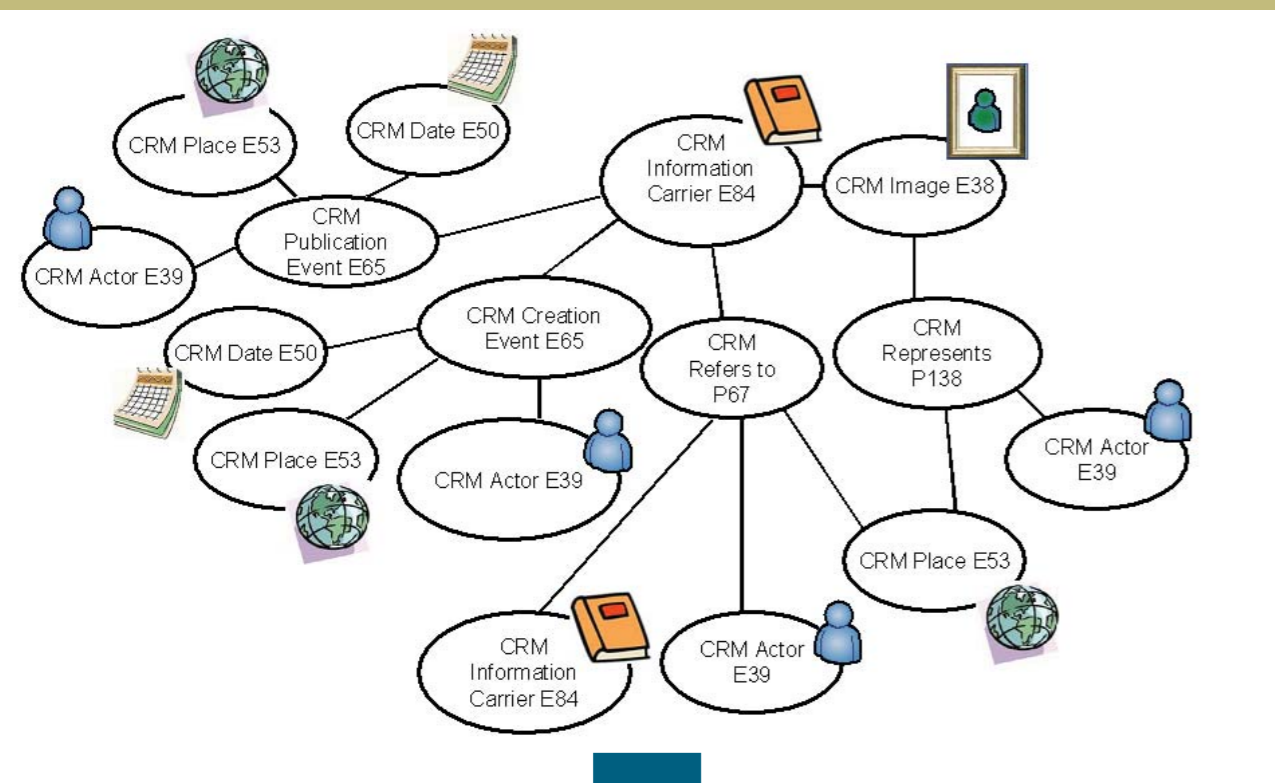
An example of some of the huge number of linkages and cross-references between texts in the collection.

Two of the texts included in the NZETC online collection are James Cowan's "New Zealand Wars" and Elsdon Best's "The Maori As He Was". Both contain drawings of the Ngati Toa leader Te Rangihaeata. Six other texts in the collection make reference to Te Rangihaeata. The portrait included in Chapter Three of Cowan's history was drawn by Charles Heaphy. There are two other sketches by Heaphy in two other texts in the collection.

Furthermore Heaphy's actions in the Taranaki War are described in A. J. Harrop's "England and the Maori Wars" and Heaphy is referred to as the source of information on traditional Maori bird-hunting spears in Best's "The Maori Volume 2". Harrop's bibliography includes texts by both Cowan and Best. There are six texts by Best in the NZETC collection, one of which was published by R. E. Owen who also published Cowan's "New Zealand Wars".
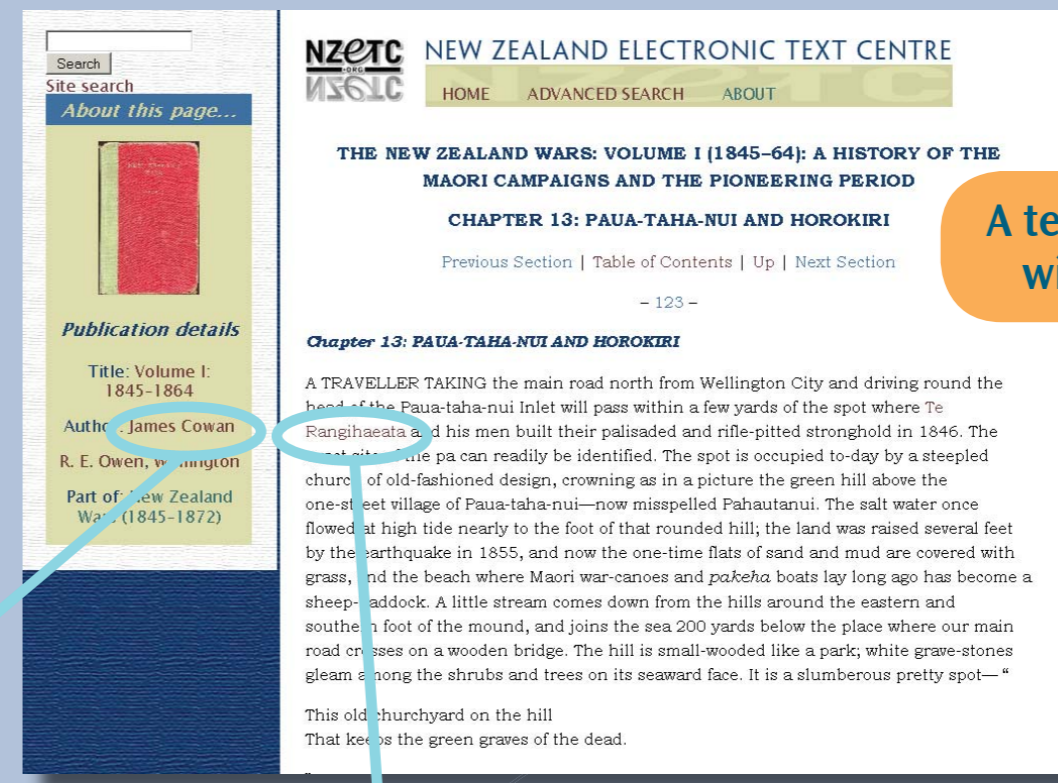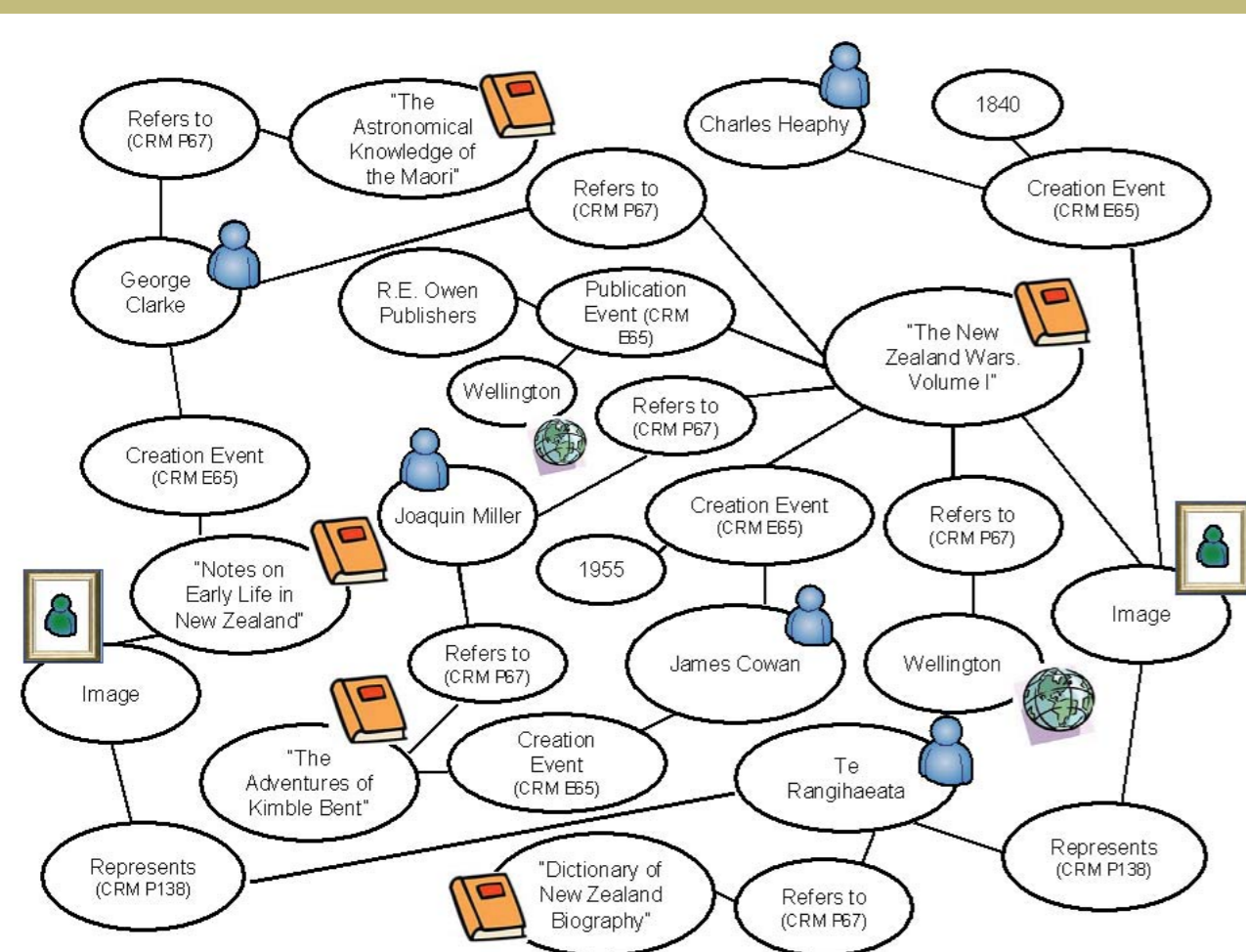
An analysis of the collection allows identification of various topics of interest. Topics in the NZETC digital library represent authors and publishers, books, chapters, images, as well as people and places mentioned in those books, chapters, and images.
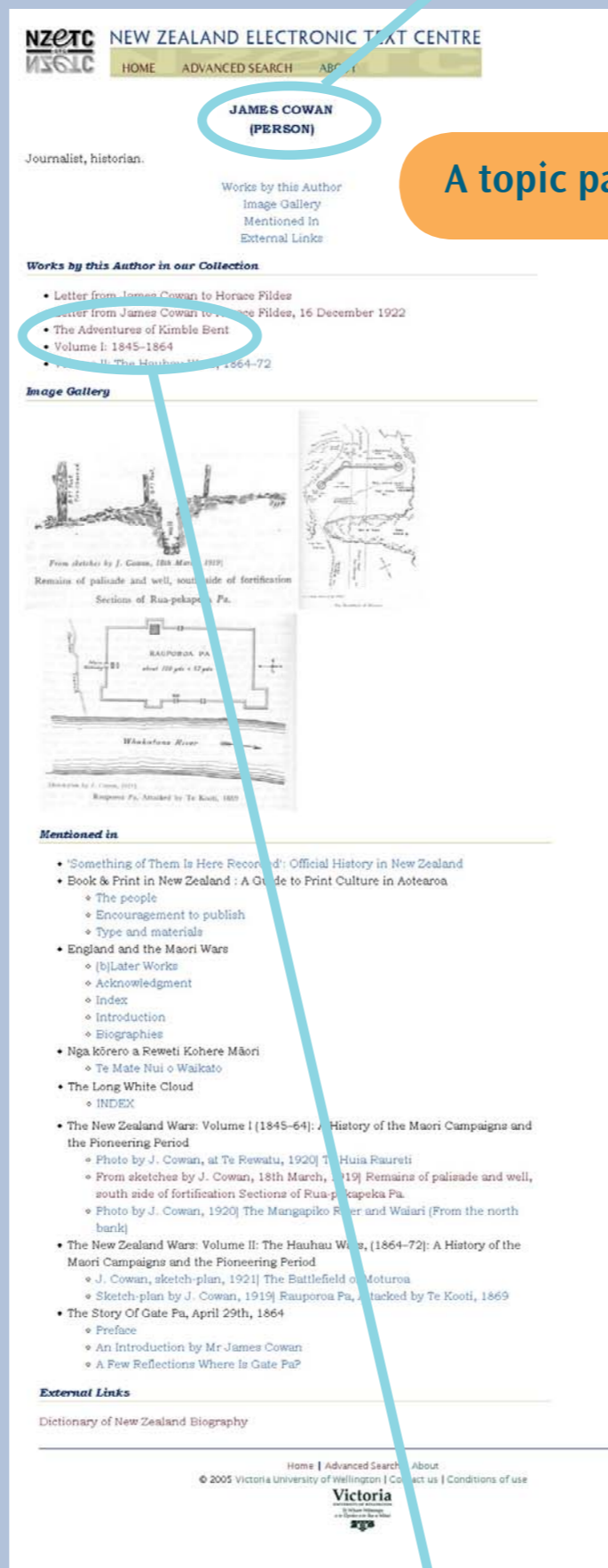
Topics of interest and the relationships between them are codified using an existing ontology, in this case the CIDOC-CRM.
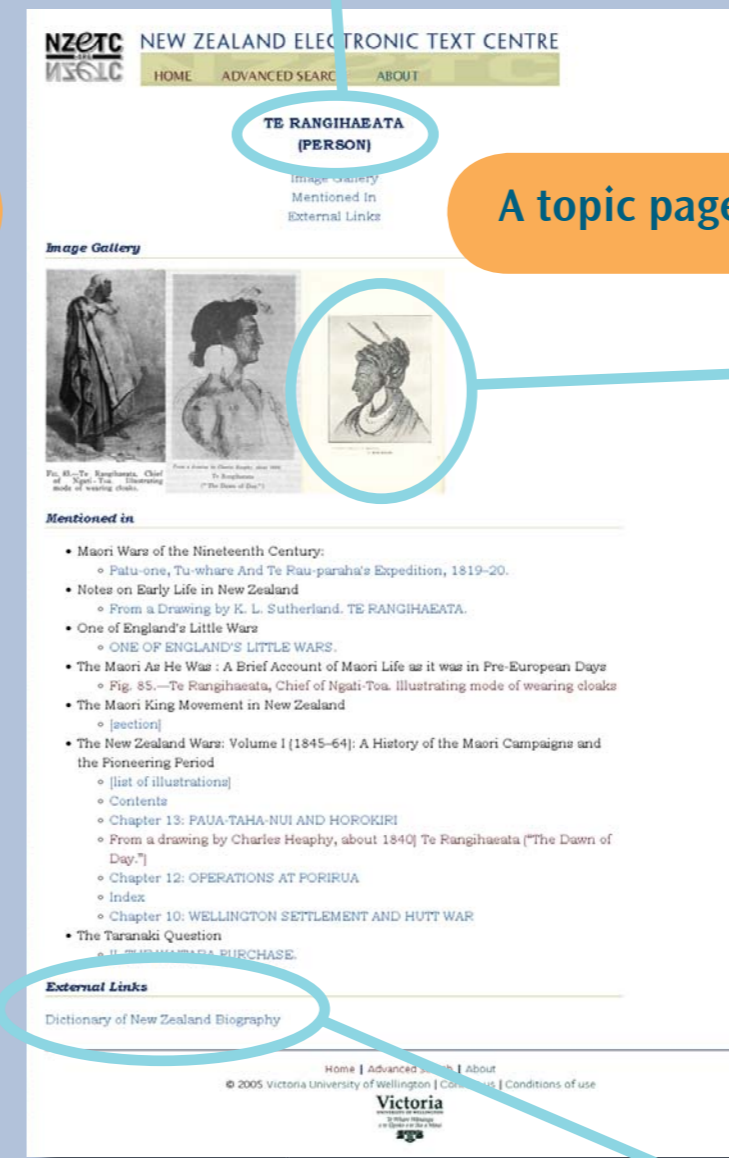
Information from the texts is harvested into this framework creating a fully connected and traversable collection.
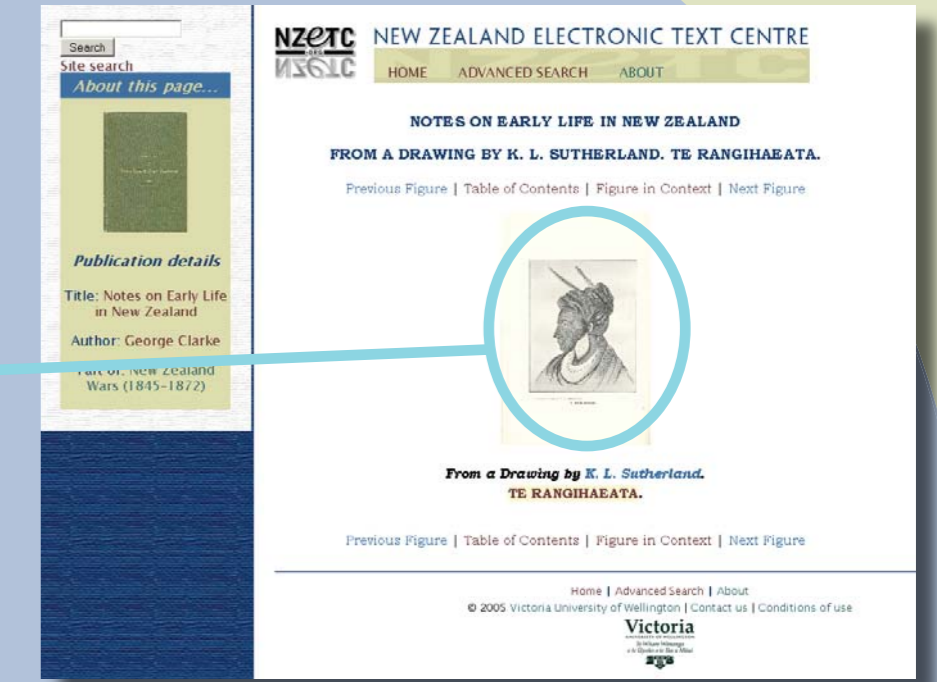

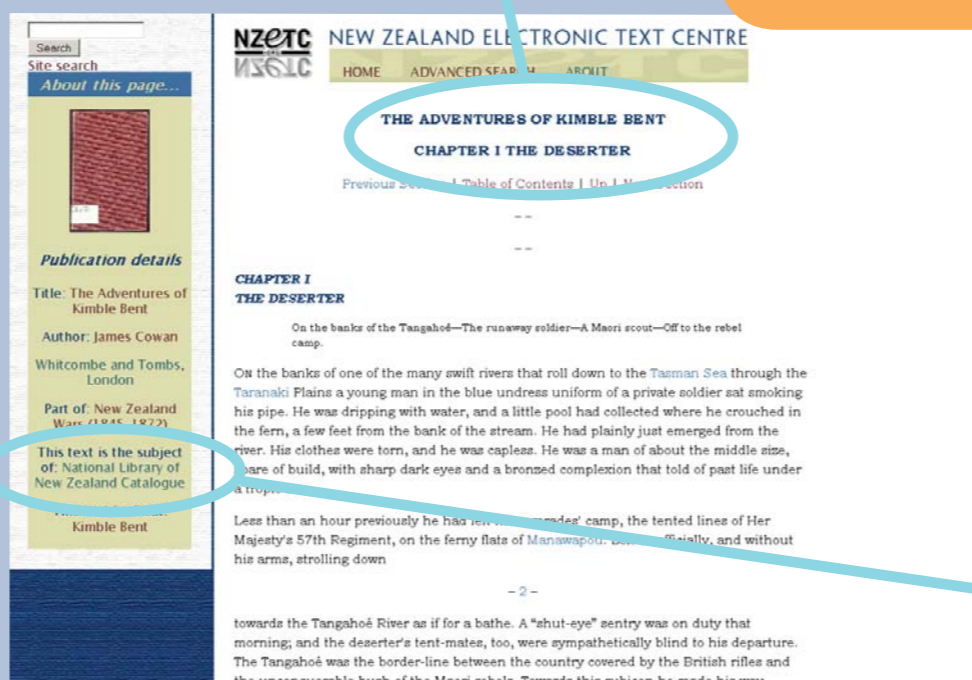A text in the collection with links to topics

A topic page

A topic page

A text in the collection

An external collection

A text in the collection

An external collection

The NZETC digital library contains over 1800 New Zealand and Pacific Island texts ranging from history to literature to language and reference. There are around 50,000 topic pages for people, organisations, places, and texts. To add a text to the collection and incorporate it into the topic map framework, the following steps are taken:

1. Printed material scanned, text transcribed through double keyboarding process.

2. Transcribed text encoded as TEI XML. Automated mark-up of references to topic of interest (ie those already identified in MADS authority file). Additional manual quality checking and markup performed where necessary.

3. Metadata automatically harvested from marked-up text into the topic map.

4. Topic Map and source texts rendered as HTML for presentation to the user, with automatic generation of hyperlinks to support sophisticated browsing.

The topic map provides a meaningful structure based on real world entities through which the user can explore the rich content in the collection and find other relevant online resources.

FUTURE WORK

- Extend topic map to include event and subject based topics using automated content analysis techniques.

- Work with other providers of other online collections to share topic identification information and thus increase linkages between collections.