Connections and Contours
Modeling Textual Data in the Online Cultural and Historical Research Environment

Andrew R. Burlingame
NELC, University of Chicago

Textual Artifacts from Ras Shamra and Ras Ibn Hani

Excavations beginning in 1928 and 1929 resulted in the discovery of ancient Ugarit. The site has given up about 2000 tablets bearing texts in the local Northwest Semitic language, Ugaritic, which was written in a previously unknown alphabetic cuneiform script. The archaeologically contextualized nature of the find makes it an invaluable source of information for study of the Late Bronze Age Levant but also demands a highly integrated approach to data modeling.

Investigative Parameters
The corpus of texts from Ras Shamra and Ras Ibn Hani represents eight different languages and a host of different textual genres. Each text is written on a tablet or object—artifacts in their own right, with unique find-spots and associated archival and architectural contexts.

The tablets and objects bearing these texts each have their own physical properties (material, dimensions), as do the texts themselves (dactus, « miroir en tablette »). People and places mentioned in these texts can be further contextualized geographically.

Input and Atomization
Textual input is based on a unicode character inventory.

Upon input, texts are automatically atomized down to the minimal epigraphic units, which are each registered as items in the database, to which attributes and links can be assigned.

Epigraphic Attributes
Damaged signs and lacunae can be indicated upon initial input of a text into the database. Indications of damage and absence are registered as metadata of the epigraphic unit concerned, to which the database responds by making attribute description options available, facilitating efficient data entry at a high level of granularity.

Epigraphy and Discourse
Texts are entered according to their epigraphic layout, but interpretations of the organization of textual content often differ from the spatial organization of the texts on the tablets. OCHRE allows the researcher to organize the epigraphic units into discourse units, which are registered as individual items to which attributes and links can be assigned. The researcher is not forced to decide whether to enter or display the text according to its material or logical organization.

Research Potential
OCHRE offers a number of tools that enhance the potential investigative and publication applications of hosted database projects like RSTI. Each lexeme in a text can be identified with a form/lexeme already in the dictionary or can be assigned a new parsing/gloss (resulting in a new dictionary entry). Personal names can be described linguistically or linked to persons in the list of ancient persons. Toponyms can be described linguistically or linked to geographical locations. These tools facilitate lexical, prosopographic, and geographic research.

Linked images or handcopies can be marked up with geometric hotspots, linking the image of each sign with its corresponding epigraphic unit in the epigraphic hierarchy. When this is done on a wide scale, palaeographic script charts can be instantly generated and organized according to the parameters specified by the researcher (e.g., by genre, date, find-spot, etc.).

Options and Attributions
RSTI/OCHRE does not restrict the researcher to a single epigraphic or discourse representation of a given text. Multiple editions can be entered and associated with their editors. Attribution remains transparent with respect to editors and database entry creators alike.

Acknowledgments
My participation in the RSTI database project has been facilitated and overseen by Dr. Miller C. Frosier, RSTI project co-director together with Professor Dennis G. Pardoe. The project is supported by the OCHRE Data Service of the Oriental Institute of the University of Chicago and has been made possible by the creative efforts of Sandra R. Schlenk (OCHRE developer) and J. David Schloen. Tablet photographs courtesy of PhoTEO, Mission de Ras Shamra.

Item-based Data Modeling
The Ras Shamra Tablet Inventory (RSTI), a database supported by the OCHRE Data Service of the Oriental Institute of the University of Chicago, is designed to curate these data digitally in a manner that facilitates such multi-parameter investigation. A flexible upper ontology allows the end user to define the parameters relevant for the investigation. Individual data-points, as well as relationships between data-points, are defined as items which property attributes can be assigned. Significantly, items can also be linked to other items in the database, and these links can be displayed together with the principal item.