I want to thank the organizers for inviting me to participate in this workshop, even though my area of interest is not the Old Assyrian period. As the co-director of the Ras Shamra Tablet Inventory database project, my area of interest is LBA Ugarit. While the nature of the data is slightly different than the Old Assyrian corpus, I struggle with the same decisions. It seems our most vexing problem is the problem of disambiguating persons who share the same name. Welcome to the study of the ancient world. I suppose the issue would be true for the modern world as well, if not for the availability of other information like street address, phone number, passport number, and social security number. We moderns have devised various systems for disambiguating in our daily lives. My phone doesn’t even need my name. It just needs a quick glimpse of my face to identify me.

But we’re left with much less information in our ancient texts. Thankfully, we have primary documentation at our disposal. From Ugarit, this primary documentation takes the form of personal correspondence, legal documents, lists of work assignments, ration lists, and other texts where personal names appear, even the famous mythological texts. The real struggle at Ugarit is the somewhat rare use of patronymics combined with the rather low instance of multiple attestations of names. Before the discovery of the library of texts from the House of Urtenu in recent years, van Soldt concluded that something like only 16% of names occur twice, with less than 10% of names occurring three times. In a few cases, we are able to establish a multi-generational lineage.

Another wrinkle in the Ras Shamra corpus: we are faced with a multi-lingual aspect to our data. The same people occur in the Ugaritic alphabetic texts and the Akkadian texts. So, we have to control for a variety of spellings of names, sometimes alphabetic, sometimes syllabic.

What I’ll be demonstrating today is a practical methodology for doing prosopography in a database environment.
Why do you find a Late Bronze Age interloper in your presence today? I’m here not only as someone who studies Ugarit, but also in my role as research database specialist at the OCHRE Data Service of the Oriental Institute. My goal today is to present a methodology that was developed for use within the OCHRE database. It’s a general methodology that has been developed to apply to a wide variety of prosopography research projects. Even though I’m working in LBA Ugarit, these principles apply to Old Assyrian. Ed is using a version of it for his database project. The Persepolis Fortification Archive project led by professor Matthew Stolper is using a version of this method to study persons mentioned in the administrative texts from Persepolis. A recently-initiated project is using this approach to study the Murashu Archive of Late Babylonian Nippur. Another project is working on the people in the Demotic economic texts from Ptolemaic Egypt. And interestingly, a slightly modified version is being used to model to social and professional world of Charles Darwin. These are all projects using the OCHRE interface and some version of what I will present today. There are a few underlying core principles shared between all these projects, but the precise application of these principles varies from project to project. I will attempt to identify the principles, then demonstrate how they apply to the Ras Shamra corpus.

**Before I go on**, I just want to be clear on these acronyms that I’m using. RSTI is the abbreviation of my research project, the Ras Shamra Tablet Inventory. OCHRE is the database platform in which RSTI exists and is the acronym for the Online Cultural and Historical Research Environment.
In summary terms, here are those principles I’ll demonstrate today. **First**, from the perspective of the database, we’ll cover the concepts of itemization and integration. We’ll see how OCHRE goes to the extreme to break down data into small and manipulable pieces, then puts them all back together in meaningfully integrated ways. **Then** we’ll see how this approach keeps separate the idea of the names list, the persons list, and the texts. The names list is essentially a glossary of names, with only one entry per lexically unique name. The persons list means to identify and disambiguate individuals, so here there is one entry per real person. The multiple instances of names in texts is a third, but very important category of data. As you will see, this approach uses the textual data as the source of information about the persons.

As Professor Larsen observed in his pre-conference thoughts on prosopography, we’re working toward identifying a biography of individuals. Of interest to me in the Ras Shamra corpus—more so than reconstructing biological family networks—is the identification of networks of power. This includes the identification of vocations and social roles more than the identification of kinship relations. Although of course, where we can identify kinship relations, we are very happy.

My approach is firmly rooted in the primary sources. In fact, as you will see, the entire process begins with the text. Any identification of a person is done so in a text. No aspect of that person is observed, unless it is observable in a text. All observations of vocations and social roles are directly linked to evidence from a text. Not only do I think this is the best methodology in terms of the data model, it also provides a mechanism for the reader to quickly validate my conclusions. This approach lays bare may data. I put all my cards on the table. Nothing is hidden behind a computer screen. It’s all there for you to click through and investigate. This is only possible because the data is highly itemized and completely integrated.

// I was heartened to read Professor Larsen’s idea of separating the index of names from the index of persons. I may have actually cheered out loud when I read it. This was one of the important breakthroughs in my own research. Once I realized I needed to keep these two separate, the remaining data model fell into place. On the one hand, we have a glossary of names. On the other hand, we have a list of the persons identified in our primary sources. The glossary of names has one entry for a name like Abdimillku. The list of persons has an entry for every disambiguated Abdimillku encountered in the corpus.
The Ras Shamra Tablet Inventory is a work-in-progress. We have no significant funding! We’ve had a few graduate student workers over the years, but the main bulk of work has been done by me. Many text editions have been supplied by my co-director Dennis Pardee.

The project is not yet at the place where a comprehensive study of names can be made. Quite frankly, we’re currently concentrating on entering texts. As you may know, there is no single or reliable repository of texts from Ugarit. The first publications are generally reliable, but need updating. Unfortunately, the various secondary collections of texts have sometimes resulted in less reliable editions. Then of course, there is a set of unpublished textual material. I’ve applied the following prosopography method to some of the texts already in the database, as a test of the principles and concepts. Some of the other OCHRE projects have been good test cases for the methodology, just to make sure there isn’t anything in the methodology that is customized for Ugarit only.
Let me give you a quick tour of the dataset in RSTI.

At the root of my methodology is the primary data. Everything begins with the tablets and texts. Any prosopography I create will be based on the textual data.

As a project, our very first task was to enter descriptive data about the objects. This data was adapted from Bordreuil and Pardee's publication La Trouvaille Épigraphique d'Ougarit, known by the abbreviation T.E.O. We have a basic description, find spot, and some bibliographic information for 5,790 objects.
As I mentioned, we’re working through the importation of texts. I won’t go into this process in detail, but OCHRE ingests a normal looking transcription. With a few exceptions, there is no need to add any special encoding. I simply feed a Word document to OCHRE and the program ingests it.
Here is a view of the text after import. From left to right, we have panes representing a description of the text, a transcription, a “vocalized” text representing something like a normalized text, then a translation, and finally a list of associated images.
Speaking of images, this was our second major task, which is now complete. The epigraphic mission to Ras Shamra had accumulated tens of thousands of tablet photos, taken over the course of many visits to the Museums in Damascus, Aleppo, as well as the Louvre. The images were imported into RSTI and linked to the associated texts. We recently also acquired photo corpora from various other researchers who generously donated their data to the project.

We have also added hand copies from various sources, published and unpublished.

Here you see an image that has been marked up with regions that represent the alphabetic Ugaritic letters. In this view of the image, you can click on a letter in the text and the letter will light up in yellow on the image, and vice versa. Click on the letter in the image to find it in the transcription.
To the basic bibliography from TEO, we are adding any bibliography we can find. Pardee has a lifetime of bibliography, some of which he has published in topical articles, much of which he simply tracked in an old FileMaker database. This information has also been integrated into RSTI.

**OCHRE provides** integration with the free online bibliography management software Zotero. I maintain a bibliography in Zotero, then simply link citations from Zotero to texts in RSTI.

**Zotero** is a very convenient tool, free to use. It has a web app version, a desktop version, and allows integration with other programs like OCHRE through means of an API.
We’re also in the process of creating a glossary of all the Ugaritic and Akkadian words in the texts, not just the proper names. We are building this glossary as we enter and analyze the texts, so it is also a work-in-progress. As I work through a text, OCHRE helps me add new attestations of words to the glossary. This will be an important aspect of the prosopography study because I’m interested in the various terms for social status, vocations, and other groups in the socioeconomic networks at Ugarit. This is the same glossary where I create my index of names.
We recently began adding GIS data to RSTI. Here you see an overview of the royal palace. Each room and courtyard in the palace is represented by a polygon in a GIS shape file. The dots represent a sort of heat map of the tablet find spots.

This may seem like an unusually broad range of data. It is. But my goal is to leverage all of this data, even the spatial layout of the find spots in analyzing the structures of power at Ugarit. I want to contextualize the textual data within this spatial layout. As for this specific spatial data, it may seem obvious that many of the texts come from the royal palace, but we’re beginning to see that we can tease out specific rooms in the palace that seem to deal with specific expression of royal power, such as conscription and rationing.
So, we come to the traditional focus of prosopography, the people. While I am interested in identifying some of the traditional types of information in the realm of prosopography—like who is sending and receiving letters—as I said, I'm more interested in modeling the social and economic organization of the kingdom of Ugarit. I'm planning on using biographies of persons as one vector of investigation.
I’ve been building a sort of mental map of the power structures represented in the texts. My goal is to test this heuristic model through prosopography. In very broad brush strokes, I see various groupings of individuals vis à vis the power structures of the kingdom. The primary actor and most powerful node is the royal palace. But I think we can observe other power centers, such as various major householders, some of the villages, and possibly vocational groups.
It is my working hypothesis that there were a few modes of interactions between these various nodes. The rules for structuring trust between individuals and groups varied from strictly kinship rules, to more contractual terms. And in between these extremes I think patronage was an operational mode of structuring trust. And at Ugarit, there seems to have been a pervasive form of royal patronage, one so embedded that it may have organized most of the kingdom.
The texts bear witness to the movement of various productive resources between the various centers. Again, my hypothesis is that royal patronage governed these transfers. In its most condensed definition, the royal patronage system at Ugarit **granted** access to productive resources in exchange for a return of value. The system was not nearly as transactional as this sounds. Instead, it was likely highly embedded in the social structure.

My goal is to tease out the observable aspects of this system using the OCHRE database and the following prosopography method.
So, how does one begin with the raw data of texts and build up a system of analysis that can address these questions?

The answer begins with a different kind of database. And I should point out that I’m not going to explain much about OCHRE today apart from how it applies to prosopography. I won’t explain much about the underlying database structure, or the computational ecosystem surrounding OCHRE. Many of you are already familiar with OCHRE, so no need to retread this group. But please feel free to ask questions later.
Now we get to the database concepts at play. First, the concept of itemization in which everything is a thing. What do I mean by this cryptic phrase “everything is a thing”?

Every tablet, person, text, and even every sign, is a separate database item. This allows for maximum flexibility for creating relationships between things. That really is the essence of prosopography research.
When I say that every sign in every text is a thing, here is what I mean. In this text alphabetic Ugaritic text, for example, the first word is probably a proper name Sākinu. In the database, the s is an item. The k is an item. The n is an item. They are related to each other so that the database knows that these three letters create the name Sākinu.

Ok, don’t get lost in the details here. Let me bring it back to reality by explaining why this is necessary. This itemization is very important because it allows the glossary of names to know the exact character-by-character spelling of any name, whether Sākinu is spelling alphabetically like this, or syllabically in other texts.
So, the database understands a lot about every letter. It **knows** how to display it using Unicode, for example. But more importantly, it knows that these three letters in this exact text make up the name Sākinu. You can probably extrapolate already the power that this will give us later when we want to find all attestations of a person in all texts, regardless of how the person’s name is spelled.
Integration

❖ All data in one place
❖ Available for various types of analysis
❖ Digital publication directly from OCHRE
❖ Create archival backups

Of the texts we have entered thus far, RSTI has over 180,000 signs in the dB. Again, each one of these signs is a separate dB item. So we need to make sure these stay organized. OCHRE does all this work for us of course. One of the primary strengths of OCHRE is the integration of all project data in the same place. I’ve already hinted at the fact that RSTI has object, text, image, and GIS data fully integrated. I won’t be talking today about digital publication and archiving. But I can promote the upcoming ASOR session on OCHRE where various speakers will talk about the complete life-cycle of data in OCHRE. See you in Denver.
For our purposes today, I want to outline the data model I am using to approach a digital prosopography of Ugarit. You’ll see why it was important to preface the talk with a brief overview of Itemization and Integration. From this point onward, practically everything you see relies on these two concepts.

First, I’m going to begin with a diagram that is meant to illustrate an abstracted representation of the integration of data in RSTI. It all begins with an object, which is linked to another database item representing the text. In OCHRE we keep separate the idea of the object and the text. A text is composed of signs. I won’t get into the details here, but every sign in the text is validated against an internal writing system. You can think of this writing system as a database version of Borger’s MZL, expanded as needed. A sequence of signs makes up a word. When a word is a proper name, it is linked to a person in the index of persons. All words are documented in the glossary. Under each lemma in the glossary we identify a grammatical form and its various attested spellings. Here, a proper name would occur once as a lemma but possibly many times for each grammatical and attested form. Important also for the prosopography study is the ability to describe any of these database items with properties.
So, any of these database items can have properties that are appropriate for describing it. An object, like a tablet, can have `lat-long` coordinates to help locate the find spot. The signs in the writing system have properties to identify them in various ways. A text might have a property that identifies the script used, the language, or the genre. Items in the glossary have parse properties. The lemma can be tagged as a noun. The grammatical forms have more specific properties. Importantly, the persons can have properties that identify their kinship affiliations. The specific attestations of words in texts can have properties that identify information about that person as attested in that specific text. All of these properties can be queried to find information in the database, or to create subsets of data for analysis and display.
This extracted portion of the data model illustrates the important separation of data. In the end, we have three primary categories of data regarding personal names. We have the glossary, which serves as an index of names. It is here that we record information like etymological properties of the name elements. Is the name Semitic, Hurrian, Egyptian, etc. Sometimes I will also give a literal gloss of the name. There will be only one entry per name here. There is only one entry for Abdimilku, for example. This has no impact on the identification of real persons in the text. It's simply a glossary of all known names.

Then we have the list of persons, which identifies and disambiguates real, observed persons. We can add properties here to identify the people by gender, whether they are historical or literary.

Most of the observable properties of a real person are left to the final category, which is the instantiation of a proper name in a text. Words in the text represent specific instances of names. These link to the dictionary so the database knows that this word in the text is a proper name. The word in the text also links to a person in a list of persons. Multiple words in texts might point to the same person. That is, the many attestations of a person in the texts all point to the same person. The word in the text is where we put most of the meaningful properties. We can identify the vocational or social role evinced by the attestation of this name in this text. This is very important, because a person's role may change over time and we want to be able to identify the source of information for any given observation about a person. The same is true for one's kinship relations. Perhaps the text here says that Abdimilku is the son of the Queen. We indicate that information on the word in the text. Because all the attestations of Abdimilku in all the texts are linked to a person in a list of persons, the database can gather and display all these observations in one place. So, all of the properties from all of the instances of Abdimilku from all of the texts are shown here on the person that represents Abdimilku. We are also given text citations so we can click through and investigate those instances.
Let’s take another look at this network of data. Let’s say we have three texts. Text 01 attests an alphabetic Ugaritic spelling of the name Abdimilku. In that text we see that he has this special status called bunushu malki, or “man of the king.” And we also see in this text that he is said to be living in the capital city of Ugarit. In Text 02, we have a logo-syllabic attestation of the name and here we see that he is serving in a military capacity in the capital city. In the third text, we have another alphabetic attestation of the name. But here we see that Abdimilku is a field owner in one of the outlying villages. Hmm, this may be a different Abdimilku. All three of these attestations of the name will link to a glossary. The Ugaritic names link to an Ugaritic glossary. The Akkadian words are documented in an Akkadian glossary. These two words are cross-referenced to make it easy to see related names. But next comes the fun part. And I’ve already tipped my hand here. I think the first two instances of Abdimilku refer to the same person. So these two words link to the first example of Abdimilku. This third guy, the one who is a field owner way out in the country, for now I am going to keep him separate as Abdimilku #2.
How do we accomplish these tasks in OCHRE? I want to emphasize that the database tools we created do not intend to replace scholarship. This method still requires human intervention. It still validates and promotes traditional philology practices. At points, the dB behaves intelligently and may save the researcher some time, but it never obviates the role of the scholar. I make this point explicitly because there is an ongoing discussion in Digital Humanities circles about the negative role technology plays in replacing the traditional role of the scholar. In this method, there is still a role for traditional scholarship. There are a lot of decisions to be made that require knowledge beyond the computer.

Returning then to the workflow, the prosopography task is one of a three-part process. The three parts are: (1) using the glossary to identify words in texts; adding new words as needed (this includes proper names), (2) using the index of persons to identify people in the texts, and (3) identifying places in the text. The first step is a requirement. Steps two and three are independent of each other and optional.

Let’s take a look at these processes in OCHRE.
OCHRE provides a workflow wizard that walks through every word in the text. Where OCHRE knows the word already, it looks it up in the glossary and identifies it. In cases where it does not yet know the word or cannot decide between homographic words, the wizard pauses and asks for human interaction.

The user can add new words or new grammatical forms on the fly in this process.
Here's a specific example. OCHRE did not yet know this word. So, the wizard pauses and gives me a chance to add it to the glossary. I can also add grammatical parse properties to the word. This is the set of properties I use in my Ras Shamra project. Each project is free to define its own vocabulary of parse properties.
Here’s an example where OCHRE did not yet know a proper name. The wizard pauses and gives me a chance to add the name to the glossary. Remember, each name is added only once to the glossary. Once I’ve added all the words to the glossary, I can move on to the next wizard, which helps me identify prosopography style information.
The second workflow wizard aids in identifying persons in the text. Like the lexicography wizard, the prosopography wizard iterates over the words in the text and stops at the first word it recognizes as a personal name. It knows a word is a proper name based on the properties I added in the previous step. Now I'm going to try to connect the word in the text to the person in my index of persons.

When the wizard finds a personal name in the text, it performs a query of all known names. The prosopography wizard returns a list of matching persons, allowing me to select the correct person. This is the chance to disambiguate between person 01 and person 02. If no matching person is found, I can add a new person on the fly. Once a link is created between the person and the word in the text, I can add additional properties to characterize the familial relationship, vocation, or other information about the person in the text.
Remember I said that the database collects all this accumulated data and displays it in a view of the person. Here’s an example of that view. Here we have a famous queen of Ugarit, Tharriyelli. Some of these properties were added to the Person, such as her social category, queen. But most of these properties are gathered from the instances of her name in texts.

Once enough names are identified and associated with persons, and enough relationships are defined, this data serves as the basis for social network analysis. Who is related to whom in real or fictive kinship relationships? Who provides goods, services, or information? And what sort of hierarchical or other relationships of power can be detected in these connections?
Even though we’re still entering texts, I am confident that the method will bear fruit. Here is an extremely rudimentary visualization of the senders and recipients of texts. I produced this very early in the process of developing this method. This is simply a diagram of who is sending letters to whom. No surprise that the queen receives a lot of letters.

The point with a visualization like this is that any one of these items can be clicked to navigate to the text or the person and investigate further.
Because the data in OCHRE is stored in very small pieces, it can be cobbled together for export in just about any format needed for use in external analytical programs. We're not going to reinvent various programs like Gephi in the OCHRE platform. But we do support the use of OCHRE data to feed these programs.

We have put a lot of effort into the integration of the R statistical platform and OCHRE. Projects can now export OCHRE data in a variety of formats that can be manipulated in R. We've also built a system in which an instance of the R program running on a server can pull data directly from an OCHRE project and return the results right to your computer.
We began with these principles: itemization and integration of data, then the creation of a name list, person list, both of which link to instances of names in texts.
In this method, we want to integrate all types of data. We want no form of data to exist in another database or spreadsheet. Some of this data is of primary interest to the act of creating a prosopography. Other sets of data are secondary, but still very helpful to have at our disposal. And finally, in this method we rely on the dB to help us manage a set of data that otherwise would be too large to manage. To state this another way, the dB allows us to be extremely ambitious about our project. We need not exclude anything that could be useful. The dB has a place for everything and keeps everything in its place.
Working in the OCHRE Database

from texts to prosopography