ISSN: 2473-3857
v2 n4035

Optionality of ERD Relationships: Project for the Introduction to Database Course

Sarah L. Klammer Kruse
sarah.kruse@mnsu.edu

Michael G. Wells
michael.wells@mnsu.edu

Computer & Information Sciences
Minnesota State University, Mankato
Mankato, MN 56001, USA

Abstract

This article presents a scenario designed to challenge students in an introductory database course to design an entity relationship diagram for countries and their consulates. The scenario emphasizes the optionality of the relationships, or relationship participation. The scenario has been used successfully as a class activity in a collaborative learning format where students work in pairs. The instructor provides clarification of business rules and facilitates periodic class discussion to guide critical thinking and data model development. The suggested solution and table definitions can be found in the appendices.

Keywords: Entity relationship diagram, ERD, data modeling, Optionality, Project-based learning, Relationship participation

1. INTRODUCTION

Data is a vital resource in companies today. Students preparing for careers in information systems are expected to have skills in data retrieval via queries and in creating conceptual, logical and physical designs for how the data will be stored. Generally students are introduced to these concepts in an introductory database course.

Database design begins with understanding user requirements, or business rules, then creating a visual diagram to represent how data will be structured to comply with these requirements. One of the most commonly used data modeling techniques is the Entity Relationship Diagram (ERD) (Thompson & Sward, 2005).

Various data modeling notations (i.e. Bachman, Chen, Martin) are used today to help students understand concepts such as entities, relationships, and cardinality (Pond, Polak, & Stutz, 2005/2006). A widely accepted industry standard uses Crow's Foot notation to represent relationships in an ERD (Hitchman, 2002). Notation detail is very important for data modelers to understand the business and modeling concepts (Hitchman, 2002). When Martin's standard was added to the Crow's Foot notation, it clearly allowed for optionality to be included for both directions of an ERD's bi-directional relationships (Hitchman, 2002).

While data modeling is a fundamental component of the learning objectives for introductory students, it has been found to be challenging, particularly in modeling relationships between entities (Batra & Antony, 1994). Students need opportunities to apply concepts to hands-on practice examples that simulate problems they may see in the real business world. However, it

ISSN: 2473-3857
v2 n4035

is not always easy for instructors to find sufficiently challenging activities that emphasize specific aspects of data modeling.

This paper presents an assignment used to challenge students in applying the concept of entity participation, or relationship optionality, when designing a data model. The assignment is used as an in-class learning experience where students alternate between independent work with a peer, and instructor-facilitated class discussion, to understand business rules and create a corresponding ERD.

2. CLASS BACKGROUND

Similar to the experience of other course instructors, students in our Introduction to Database course must work hard to understand data modeling. The relationship optionality concept continues to be one of the most difficult for students to fully understand. A number of hands-on activities are used during class time to help students learn concepts and apply this new knowledge. The assignment outlined in this paper is one of these in-class activities specifically designed to reinforce the understanding of optionality. It is not the students' first data modeling assignment, so some basic aspects of data modeling are assumed to be familiar to students.

Our course meets two days each week for a one hour and 50 minute block of time. A cooperative learning format encourages students to collaborate as they work to find solutions to scenarios presented and helps build teamwork and metacognitive skills allowing them to problem solve at a higher level than when working alone (Millis, 2010).

Ryan, Bordoloi, and Harrison (2000) define cooperative learning as an "approach that uses a small groups of individuals who work together on a collective task to maximize their own and each other's learning" (p. 10). During this class activity, students work in pairs. We have observed that working with a partner is an advantage when learning these concepts. Students are less likely to be passive listeners sitting back and uninvolved because they are accountable to a partner and often partners teach each other concepts (Barkley, Cross and Major, 2014).

4. LEARNING NEED

After searching countless textbooks and examples online, we could not find any that were particularly good at explaining the optionality, or relationship participation, concept to beginning students. So, we developed our own problem.

The optionality of a relationship can be either optional or mandatory. In other words, and entity either may (optional), or must (mandatory), participate in the relationship. Because relationships are bi-directional, there are two optionalities defined on each relationship. Optional participation means that an occurrence of one entity does not require a corresponding entity occurrence for that relationship. This is depicted by drawing a small circle on the side of the optional entity. Mandatory participation requires participation by both entities in the relationship and is depicted by a small hash mark (Coronel and Morris, 2015). See Appendix A for examples of these notations.

Beginning data modelers are challenged to learn conceptual skills about data and its properties, as well as how to graphically represent the way data is related. Ryan, Bordoloi, and Harrison (2000) consider conceptual database design to be one of the most complex skills to learn. Batra and Antony (1994) noted that modeling relationships is difficult for novices.

Two particularly challenging aspects of data modeling curriculum we have observed include resolving many-to-many relationships, and determining if a relationship is optional or mandatory.

The assignment presented in this paper gives students an opportunity to apply the concept of optionality to a number of relationships based on stated, or discussed, business needs. In addition, when creating the original model, students must resolve a many-to-many relationship. Discussion during this exercise also reinforces concepts of primary and foreign keys, strong vs. weak entities and relationships, and composite keys since students in our Introduction to Database course have completed other basic data modeling exercises earlier in the semester.

5. ASSIGNMENT

The assignment named "Consulate Problem" is targeted at increasing student understanding of relationship optionality. Students are given a set of business rules and asked to work with one

ISSN: 2473-3857
v2 n4035

other student to create the ERD during the class period. Students are encouraged to ask clarifying questions to better understand the business rules of the scenario as they create their data model. In the past, assignments with less ambiguity have been assigned. We have found students obtain a deeper understanding if the assignments more closely simulate a real world environment where problem statements are somewhat nebulous and students ask clarifying questions to understand the business rules.

General Instructions
Students are given the following instructions to begin the assignment:

Design a data model involving cities, the countries they are in, and "consulates." You will need to ask clarification questions to resolve any ambiguities. Be sure your model handles all of the following business rules.

Business Rules
The business rules for the "Consulate Problem" are listed below. There is some ambiguity intentionally built into this example. The business rules are purposely authored in this way to simulate a real world experience with users (Wiegers and Beatty, 2013). Later discussion explains how some ambiguity is handled. As a note when teaching this example, many of the facts stated are for teaching purposes and meant to stress concepts while creating the data model. It is up to the course instructor whether they want to "fact check" each business rule as the global political climate can change over time.

1. Each city is in a unique country. Data about cities include the name of the city and its population. Names of cities are unique within countries, but two cities in different countries may have the same name (and possibly even the same population).

2. Information about a country includes its name and its head of state. You may assume no two countries have the same name.

3. Information about a consulate includes its name, which is unique (e.g., "U.S. Consulate General Toronto" and its street address (which is unique within a city, but might not be unique among all consulates in different cities).

4. Cities, and the country where each city is located, are related by a relationship labeled *In*.

5. Consulates and the city in which each is located are related by a relationship labeled *Located in*.

6. There is no direct relationship between consulate and country. Instead, each country is represented by a consul who works within a consulate. Data stored about each consul includes name and the office assigned. (A consul is only assigned one office.) Occasionally, a consul will represent more than one country. For example, country A may not have diplomatic relations with country B, so country A will ask country C to represent A's interests in country B at their own consulates. In this situation, the consul for country A is actually a citizen of country C, and typically, is not the same as person hired as the consul for country C. Both of these consuls, who are citizens of country C, would have an office at the same consulate.

7. While you might assume that a country would have, at most, one consulate in a given city, this is not always the case. For example, the US retains two consulates in Paris, France.

Data Model Requirements
The students are required to perform the following tasks using a diagramming tool. In the past we have used Microsoft Visio to create the ERD. Recently, we have switched to LucidChart for the diagrams. The tool is not really important, as long as it can perform the following assignment requirements.

1. **Name** the entities

2. **Identify** a primary key for every entity and foreign key(s) where necessary

3. **Create** correct relationship cardinality, i.e., 1:1 or 1:M

4. **Name/Label** the relationships in the ERD

5. **Define** the relationship optionality

ISSN: 2473-3857
v2 n4035

6. **Define** entity attributes using an entity prefix concept and define basic metadata for every attribute (data type and size)

7. **Bold** entity attributes that are required

Definitions
Typically, it takes 10-15 minutes for students to read the business rules and begin to understand them. One of the first questions asked will be, "what is the difference between a consul and a consulate." Acting as the business domain expert, the instructor can explain the difference. "A consul is a person. A consulate is a building."

Primary Key Definition
After some basic understanding, each group will likely come up with the following entities: Consul, Consulate, Country, and City. This is a great opportunity to ensure they have a good understanding of how to define a primary key (PK). When defining the PK for the Country entity, some students will suggest that the Country-Name attribute would be a good choice for the PK since the business rules indicate that "you may assume no two countries have the same name."

However, the instructor can point out that on occasion, country names can change. The country of Burma is now named Myanmar. If name is the PK, a name change would be problematic because then the PK would also need to change. Subsequently, any occurrence of the Foreign Key (FK) using "Burma" would also need to be changed. The choice of a unique, short, numeric PK for the Country entity would be a better choice since it is unlikely to change over time. If a country changed its name in this scenario, only the Country-Name attribute would need to be updated. An "auto-generated" PK would be a better choice, such as implementing a sequence in Oracle.

Once all PKs have been defined for Consul, Consulate, Country, and City entities, students can be asked to define the relationships between these entities.

Discussion about Ambiguities
A general discussion should take place after students have been given some time to try and create relationships that reflect the business rules on their own. The ambiguity built into the business rules simulate the ambiguity encountered during real world interviewing (Wiegers and Beatty, 2013), yet beginning students may be unable to fully analyze the unstated details of the business rules.

The discussion should emphasize the following clarifications:

"Who belongs in the Consul entity?" It should be noted that this data model is not a world census. Only people currently employed as a consul should be entered into this table. Other people are outside the scope of this problem.

"Can anyone name the countries the United States does not have diplomatic relations with?" At the time this article was authored, according to the New York Times, there are only three countries the United States does not have diplomatic relations with: Bhutan, Iran, and North Korea (Schiavenza, 2015). This answer can be easily rechecked before delivering a lecture. This fact will become more important later in the discussion when the Citizenship Problem is discussed. It further simulates real world interviews if all the relevant information for a specific problem is not given at the same time.

"Can anyone name the current countries that made up the former Soviet Union?" This is a fun exercise to see how many countries students can name. The answer can be verified by Wikipedia, but at the time of publication were: Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan. Naming all the countries is trivial, but may promote some discussion. The important point to stress for this example is sometimes one or more of these countries get together to hire a single consul to represent all of their interests in a country.

"Can anyone name a U.S. city that has a consulate in it?" Cities like New York, Los Angeles, Chicago, Atlanta, Miami and San Francisco will likely come up in discussion. This should tip off students that a country can have more than one city in it that contains a consulate. An important follow up question is, "How many consulates are there in New York City?" There are more than 100 consulates in New York City. Again, this is a clue that there can be more than one consulate in a city.

Many-To-Many Relationship
Most students quickly realize there is a many-to-many relationship between the Consul and Country entities. This is a good opportunity to review the problems inherent to many-to-many

relationships and how these can be avoided by creating a bridge or associative entity (Coronel & Morris, 2015). Knowing that many-to-many relationships cannot be implemented in a relational data schema without creating problematic data redundancy, students will insert a bridge entity entitled, "Represents."

Once the bridge entity is created, there is an opportunity to reinforce the concepts of a primary key selection and strong vs. weak relationships. The "Represents" entity could be diagrammed with its own unique key and include the Consul_ID and Country_ID attributes as two FKs. With a unique key, the relationships connecting the bridge entity would be considered weak. A weak or existence-dependent entity "can exist in the database only when it is associated with another related entity occurrence" (Coronel & Morris, 2015, p. 123). In the solution provided, we have defined the relationships as strong and created a composite PK for the Represents relationship. This prevents multiple records containing the same combination of Consul_ID and Country_ID.

Again, leaving out the details about using a bridge entity and choosing its primary key in the original instructions allows students an opportunity to work out a solution. Students often find as they solve one business rule problem, it may create another problem that needs to be resolved. Since there is no means for automatic feedback to test conceptual designs (Batra & Antony, 1994), it is important for the instructor to provide periodic feedback. In our course, the instructor observes students' work in-progress and engages students in class discussion at the appropriate times to prevent frustration and take advantage of teachable moments.

Teaching ERD Scope

Another important concept for students to understand is the scope of an ERD (Wiegers and Beatty, 2013). What should be in the model and what should be out? Each of the following scenarios are presented and students are asked to check their data model, and adapt if necessary, to correctly handle each situation. The scenarios are designed to add a little fun to the evaluation and review of the data models. Refer to Appendix A: "Consulate Problem" ERD for a graphical representation of each identified problem.

"Good Thunder" problem

Good Thunder, Minnesota is a small town just a few miles from our university. Good Thunder has all the small town essentials: post office, fire department, restaurant and bar, church, etc. However, it is obvious that Good Thunder would not have a consulate in it. Should Good Thunder be entered into this data model? The answer should be "No." Only cities that have consulate(s) in them should be entered. Because it is not possible to monitor all end users using this data schema, student data modelers must keep Good Thunder out of the city table by appropriately modeling their ERD.

If a students' data model has an optional optionality on the "Located in" relationship nearest the Consulate entity, their model has the "Good Thunder" problem. As the instructor observes students working, he/she may tell a student they have the "Good Thunder" problem, but allow students time to figure out and correct what caused this problem on their own.

"Riverboat Consulate" problem

Use an example of a building not in a city. The one used in this example is of a river that separates two cities. Ask the students what might be built there? Possibly a casino on a riverboat. Why would that happen? Should we allow a consulate to be built outside of the bounds of any city? Not for this example.

If a students' data model has an optional optionality on the "Located in" relationship nearest the Consulate entity, then it has the "Riverboat Consulate" problem.

"Nomadic Consul" problem

As business rules are introduced, mention that a consul's name and assigned office must be recorded. Through prior questions and discussion, it would have been noted that the data model tracks only where a consul's office is located and that each consul only has one. The data model does not track everywhere a consul goes during his/her work day. Each consul is assigned a single office upon being hired.

If a students' data model has an optional optionality on the "Works" relationship nearest the Consulate entity, it has the "Nomadic Consul" problem. In other words, data could be entered in such a way that a consul is not assigned an office and must wander the building looking for a place to work each day.

"Government Overspending" problem

Through discussion it should be expressed to students that when a consulate is built, there will be a consul(s) working in it. A consulate should never be empty and a government should not be

constructing an entire building to house a single consul. If this is the case, the government has spent a lot of money on an unused/underused building.

If a students' data model has an optional optionality on the "Works" relationship nearest the Consul entity, then it has the "Government Overspending" problem.

"Japan" problem

For the purposes of this example, choose a country and assume it does not participate in any diplomatic relations. Japan could be considered to be that country. Ask students, "Does Japan belong in the data model?" The answer is "No." Only countries that participate in diplomatic relations, and thus have a consul(s), should be listed in the Country table.

If a students' data model has an optional optionality on the "Represents" relationship nearest the Consul entity, then it has the "Japan" problem.

"Lazy Consul" problem

Ask the class, "who would like a $100,000 per year job and not have to do anything?" Then tell them to make certain their data model doesn't allow that to happen!

If a students' data model has an optional optionality on the "Has" relationship nearest the Represents bridge entity, then this model would permit entry of a consul that does not represent any countries and they have created the "Lazy Consul" problem.

"Poor Country" problem

There are some countries that participate in diplomatic relations, but the country is either too poor or there is too much conflict within its borders to have a consulate built within its boundaries. These countries must be allowed into the model since they are still participating in diplomatic relations.

If a students' data model has a mandatory optionality on the "In" relationship nearest the City entity, then it has the "Poor Country" problem and the model prevents that country from having a consul.

"Vatican City" problem

Should cities not in a country be allowed in this data model? A good topic of discussion will be to ask if anyone can think of a city in this situation. If the students cannot come up with one, offer Vatican City, Monaco, or Singapore as a potential example.

If a students' data model has a mandatory optionality on the "In" relationship nearest the Country entity, then it has the "Vatican City" problem.

The solution is to add Vatican City as both a city AND a country. You can show this at the table level and add a FK to connect the entries. See Appendix B: Sample "Consulate Problem" Tables for an example of this data.

"Citizenship" problem

Make sure the students are aware that on April 7, 1980, the United States broke diplomatic relations with theocratic Iran, and on April 24, 1981, the Swiss Government assumed representation of U.S. interests in Tehran. The data model they develop must be able to store a consul, who is a Swiss citizen, representing the U.S.'s diplomatic relations while working in Tehran, Iran. This is an excellent opportunity to traverse the basic data model. Beginning with the Consul table, ask students how to determine which countries a consul represents. By finding a consul's ID as a FK in the Represents table, they can determine the countries he/she represents. Use the corresponding country ID as a FK in the Country table to find the country's name. In this case, United States would be returned. See Appendix B: Sample "Consulate Problem" Tables to see that consul John Smith represents the United States and Canada.

Next, ask students, "How do you determine where the consul is working?" Again, beginning with the Consul table, traverse the data model in the other direction using FKs going from Consul to Consulate to City to Country. This will return the city and country a consul works in. For this example, Tehran and Iran would be returned as the location of the consulate where a consul works. See Appendix B: Sample "Consulate Problem" Tables to see that consul John Smith works at the Embassy of Switzerland in Tehran, Iran.

That leaves the citizenship of a consul from the Citizenship Problem in question. Currently, the data model does not support that part of the business rules. How can this be solved? Ask the students to try and solve it. A possible solution will be to connect the Consul and Country entities with a second relationship that represents citizenship.

Talk about this potential solution and then try implementing this business constraint. There are NO dual citizenships allowed in this data model. Have students implement a new solution by adding a citizenship attribute to the Consul entity. See Appendix B: Sample "Consulate Problem" Tables to show that consul John Smith works in Tehran, Iran representing the United States and Canada, but is a citizen of Switzerland.

6. REFLECTIONS

Students completing their first database course begin with little to no knowledge about data design and have limited training time (Batra & Antony, 1994) in a classroom environment. It is important to use meaningful examples and heuristics that allow students to quickly build skills and apply new concepts within the academic time constraints.

Our experience using the "Consulate Problem" to achieve the desired learning outcomes is both heuristic and quantitative. As previously stated, we require students to work in pairs to come up a solution. Every student is expected to submit their own copy of the completed ERD to an on-line dropbox. During the most recent semester when this assignment was completed, the average score was 87.6%. Some errors in the submissions were expected since students are still considered to be novices. However, most of the errors were related to use of foreign keys, and modeling the citizenship of the consuls. Out of 67 student submissions, there were only three submissions with errors where a relationship optionality did not match a corresponding business rule.

On assignments completed after this "Consulate Problem," it was very noticeable that students asked more skillful questions during lecture and about future data modeling exercises. Because students better understood how to ask clarifying questions and model relationships correctly, performance in the optionality portion of subsequent data modeling exercises closely matched business rules. In addition, students performed well on the midterm exam questions addressing optionality.

In summary, the "Consulate Problem" is a useful assignment that gives students an opportunity to learn, and apply, concepts about relationship optionality in a data model. In just a few class hours, students successfully follow business rules to create a five-entity ERD and are prepared to incorporate these new skills into future class exercises and larger models.

7. REFERENCES

Barkley, E. F., Major, C. H., & Cross, K. P. (2014). *Collaborative learning techniques: A handbook for college faculty*. Somerset, US: Jossey-Bass.

Batra, D., & Antony, S. R. (1994). Novice errors in database design. *European Journal of Information Systems*, 3(1), 57-69.

Coronel, C., & Morris, S. (2015). *Database systems: Design, implementation, and management* (11th ed.). Boston, MA: Cengage Learning.

Hitchman, S. (2002). The details of conceptual modelling notations are important – A comparison of relationship normative language, *Communications of the Association for Information Systems*, 9(1), 167-169.

Millis, B. (Ed.). (2010). *New pedagogies and practices for teaching in higher education: Cooperative learning in higher education: Across the disciplines, across the academy*. Sterling, VA: Stylus Publishing.

Pond, A., Polak, P, & Stutz, J. (2005/2006, Winter). Evaluating the teaching effectiveness of various data modeling notations, *Journal of Computer Information Systems*, 46(2), 78-84.

Ryan, S. D., Bordoloi, B., & Harrison, D. A. (2000). Acquiring conceptual data modeling skills: The effect of cooperative learning and self-efficacy on learning outcomes. *SIGMIS Database*, 31(4), 9-24.

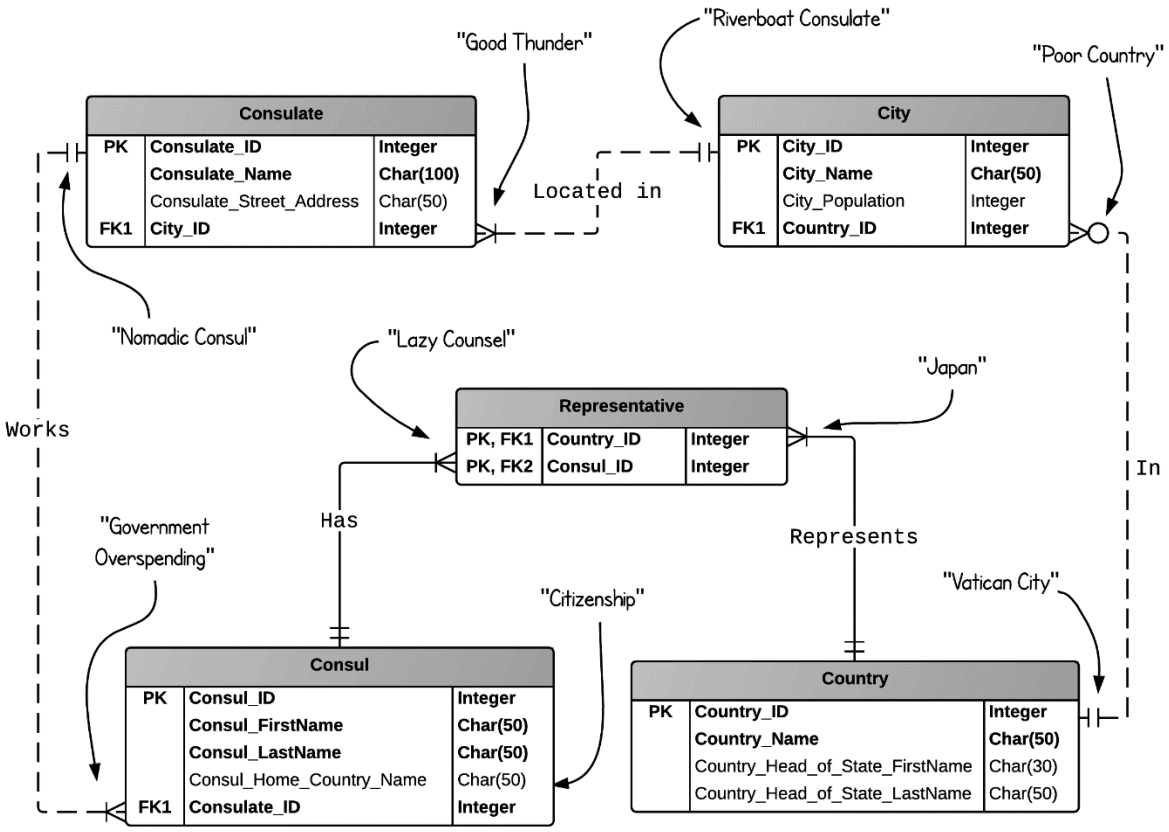
Schiavenza, M. (2015, July 1). After Cuba: The only 3 countries that have no relations with the U.S. *The Atlantic*. Retrieved from http://www.theatlantic.com/international/archive/2015/07/cuba-us-embassy-bhutan-relations/397523/

Thompson, C. B., & Sward, K. (2005). Modeling and teaching techniques for conceptual and logical relational database design, *Journal of Medical Systems*, 29(5), 513-525.

Wiegers, K., & Beatty, J. (2013). *Software Requirements* (3rd ed.). Redmond, WA: Microsoft Press.

ISSN: 2473-3857
v2 n4035

APPENDIX A: "CONSULATE PROBLEM" ERD



ISSN: 2473-3857
v2 n4035

APPENDIX B: SAMPLE "CONSULATE PROBLEM" TABLES

Table name: CONSULATE

Consulate_ID	Consulate_Name	Consulate_Street_Address	City_ID
3	Embassy of Switzerland	235 Valiasr Street	92

Table name: CITY

City_ID	City_Name	City_Population	Country_ID
11	Vatican City	451	4
92	Tehran	8,154,345	93

Table name: COUNTRY

Country_ID	Country_Name	Country_ Head_of_State_FirstName	Country_ Head_of_State_LastName
4	Vatican City	Giuseppe	Bertello
77	United States	Barack	Obama
111	Canada	Justin	Trudeau
93	Iran	Hassan	Rouhani

Table name: REPRESENTS

Consul_ID	Country_ID
414	77
414	111

Table name: CONSUL

Consul_ID	Consul_ FirstName	Consul_ LastName	Consul_ Home_Country_Name	Consulate_ID
414	John	Smith	Switzerland	3