US 20130346111A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2013/0346111 A1

(43) Pub. Date: Dec. 26, 2013

Goodgame

(54) SYSTEMS AND METHODS FOR SUBJECT IDENTIFICATION (ID) MODELING

(71) Applicant: **Joseph William Charles Goodgame**, Zionsville, IN (US)

(72) Inventor: **Joseph William Charles Goodgame**, Zionsville, IN (US)

(73) Assignee: **Quintiles Transnational Corporation**, Durham, NC (US)

(21) Appl. No.: **13/925,187**

(22) Filed: **Jun. 24, 2013**

Related U.S. Application Data

(60) Provisional application No. 61/663,292, filed on Jun. 22, 2012, provisional application No. 61/663,057, filed on Jun. 22, 2012, provisional application No. 61/663,299, filed on Jun. 22, 2012, provisional application No. 61/663,398, filed on Jun. 22, 2012, provisional application No. 61/663,219, filed on Jun. 22, 2012, provisional application No. 61/663,357, filed on Jun. 22, 2012, provisional application No. 61/663,216, filed on Jun. 22, 2012.

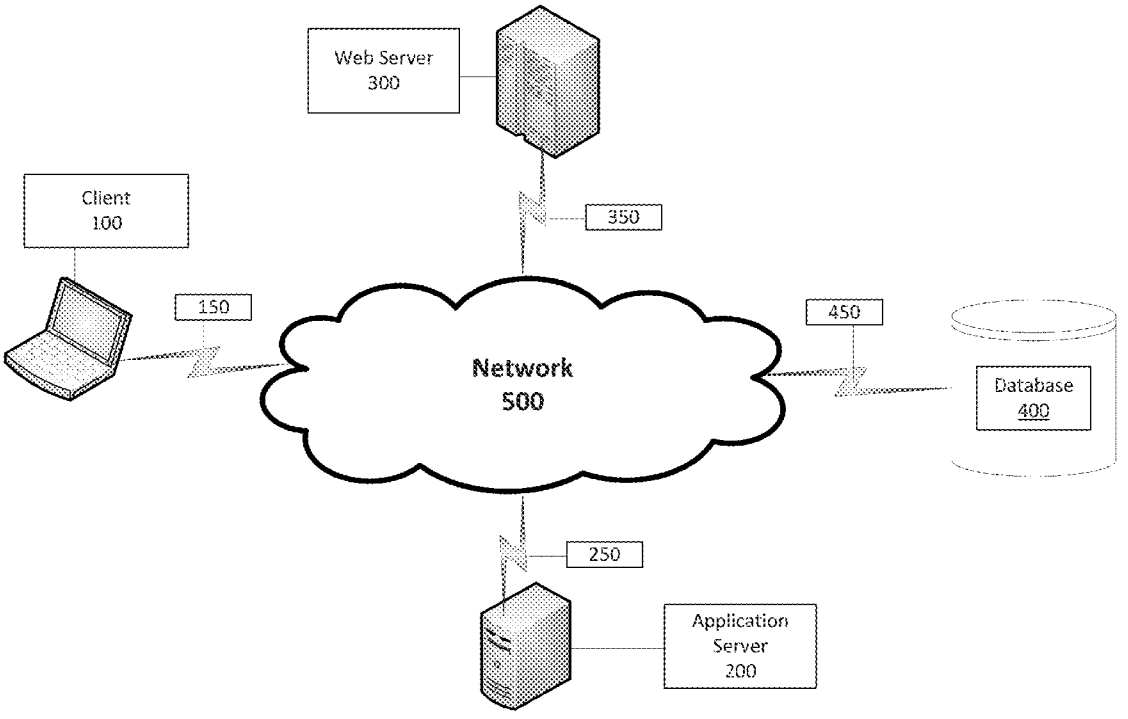
Publication Classification

(51) **Int. Cl.**
G06F 19/00 (2006.01)

(52) **U.S. Cl.**
CPC *G06F 19/363* (2013.01)
USPC **705/3**

(57) **ABSTRACT**

Systems and methods for subject identification (ID) modeling are disclosed. A subject identification may be associated with information contained in one or more core domains such as a patient domain, a country domain, and/or an investigator domain. The domains can be generically designed such that data sources that are unknown at the time the domains are created can be managed. In this way, using generic structures that support the domains, data sources can be added and/or updated as additional information and/or data sources become available. Using various graphical user interfaces, a user can dynamically associate patient criteria, country criteria, investigator criteria, and/or other information with subject identifications. A subject identification may be associated with a specified capture date. Information contained in the various domains may be filtered such that only information contained in the domain on or before the capture date is available for the subject identification.



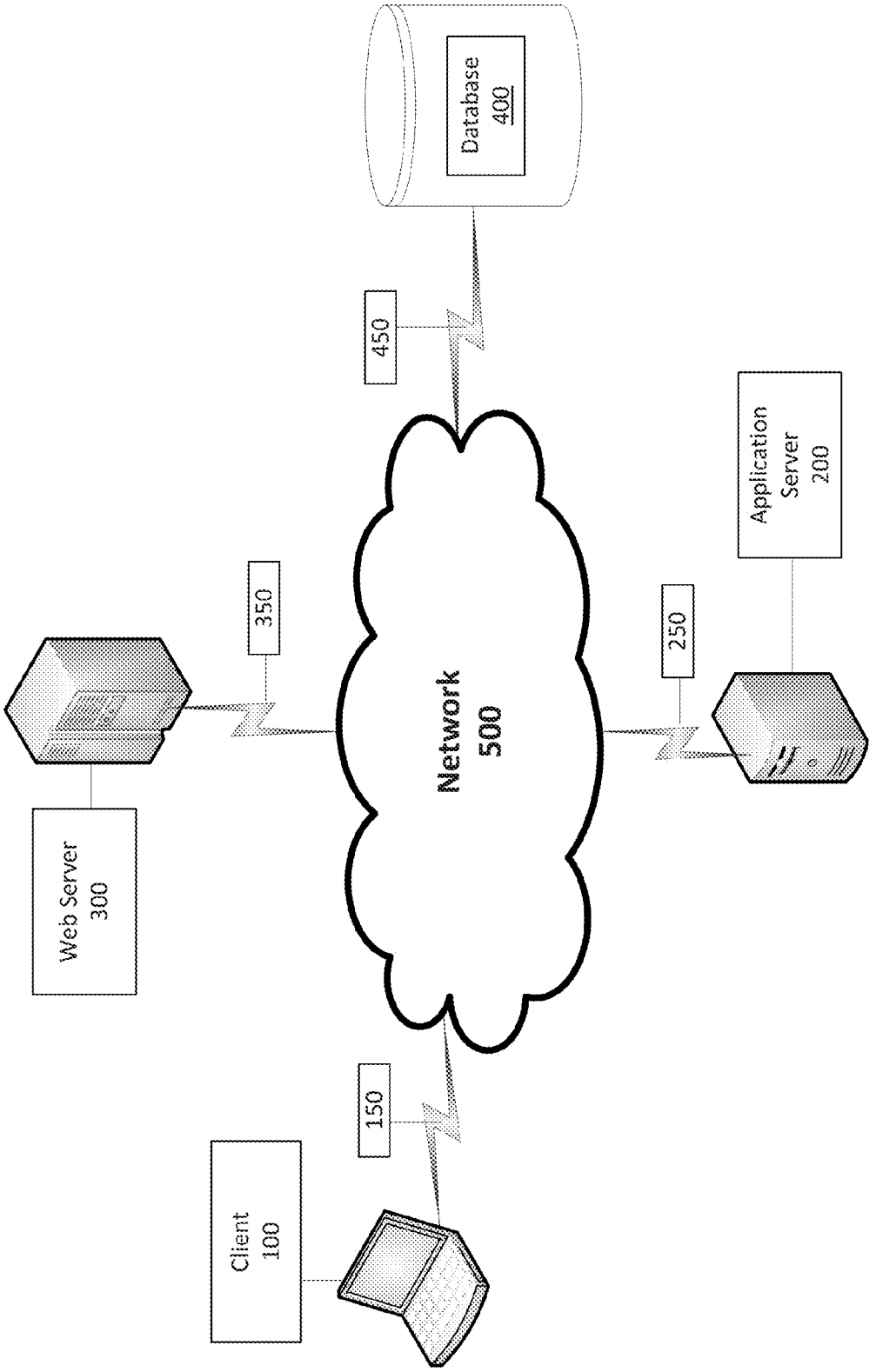

Database
400
450
Application
Server
200
350
250
Network
500
Web Server
300
150
Client
100

FIG. 1

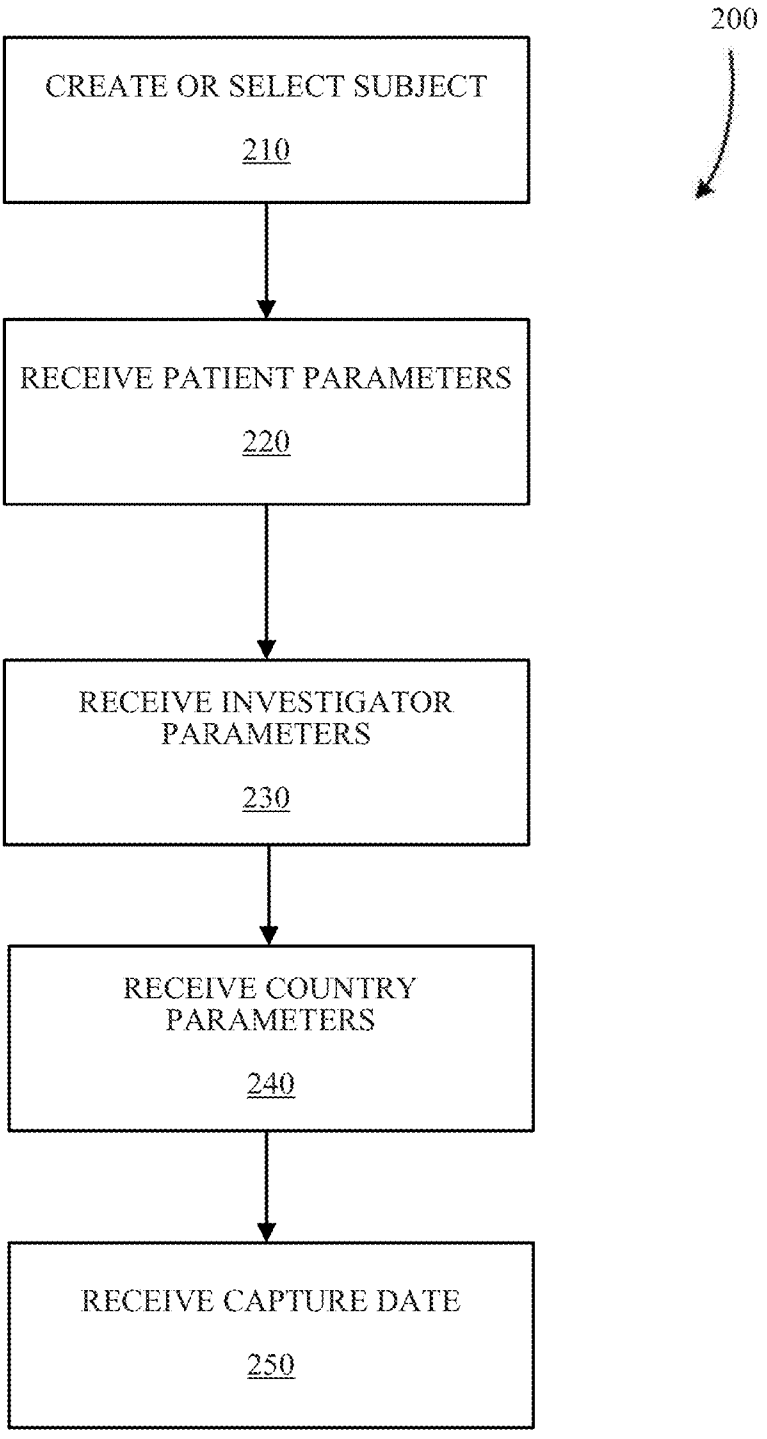

200
CREATE OR SELECT SUBJECT
210
RECEIVE PATIENT PARAMETERS
220
RECEIVE INVESTIGATOR PARAMETERS
230
RECEIVE COUNTRY PARAMETERS
240
RECEIVE CAPTURE DATE
250

FIG. 2

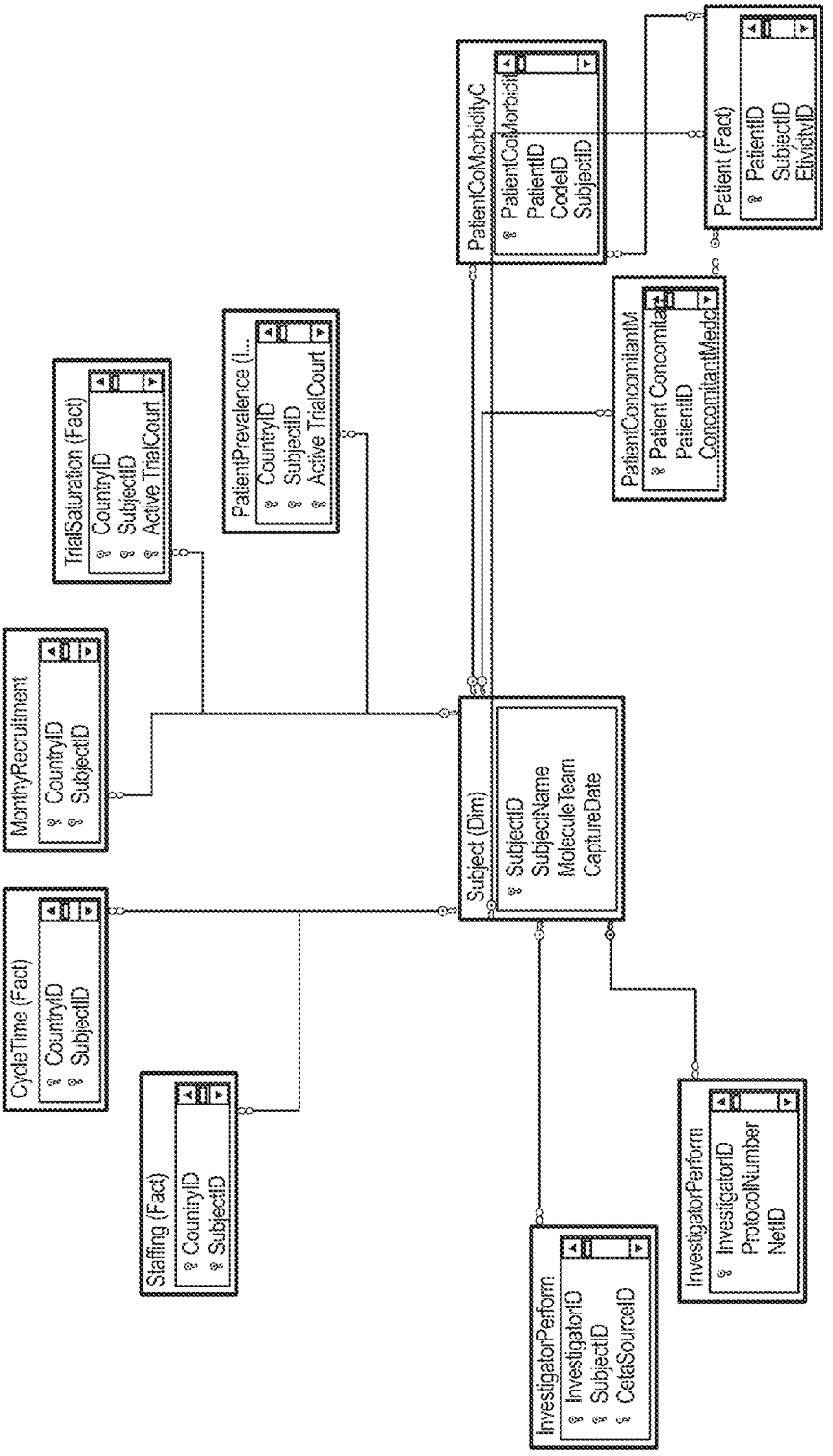

Subject (Dim)
SubjectID
SubjectName
MoleculeTeam
CaptureDate
MonthlyRecruitment
CountryID
SubjectID
TrialSaturation (Fact)
CountryID
SubjectID
Active TrialCount
PatientPrevalence (I...
CountryID
SubjectID
Active TrialCount
PatientCoMorbidityC
PatientID
CodeID
SubjectID
Patient (Fact)
PatientID
SubjectID
EthnicityID
PatientConcomitantM
Patient Concomitan
PatientID
ConcomitantMedcl
CycleTime (Fact)
CountryID
SubjectID
Staffing (Fact)
CountryID
SubjectID
InvestigatorPerform
InvestigatorID
SubjectID
CeltaSourceID
InvestigatorPerform
InvestigatorID
ProtocolNumber
NetID

FIG. 3

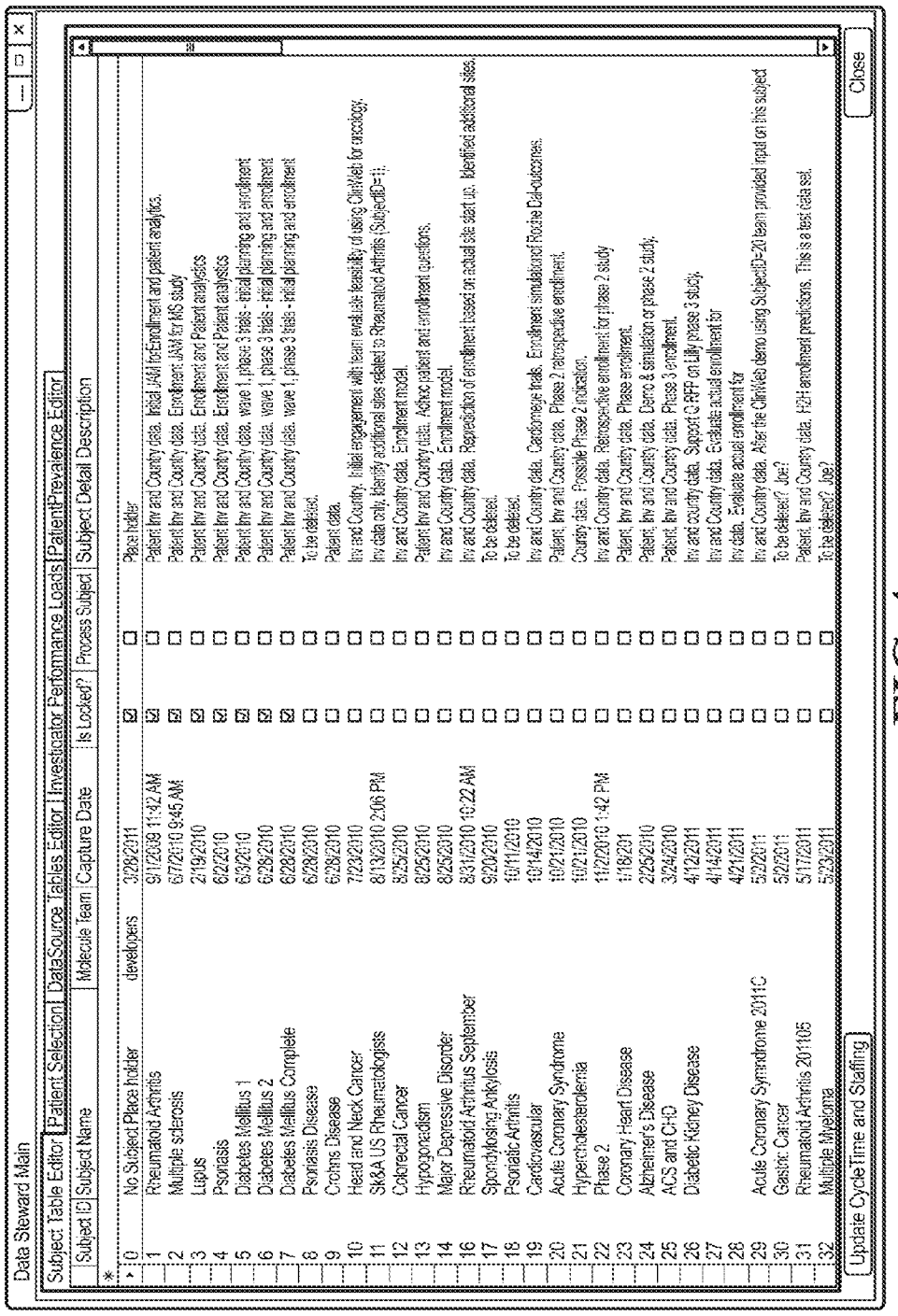

Data Steward Main
Subject Table Editor | Patient Selection | DataSource Tables Editor | Investigator Performance Loads | PatientPrevalence Editor
Subject ID | Subject Name | Molecule Team | Capture Date | Is Locked? | Process Subject | Subject Detail Description
No Subject Place holder developers 3/28/2011 Place holder
1 Rheumatoid Arthritis 9/12/2038 11:42 AM Patient, Inv and Country data. Initial JAM forEnrollment and patient analytics.
2 Multiple sclerosis 6/7/2010 9:45 AM Patient, Inv and Country data. Enrollment JAM for MS study
3 Lupus 2/19/2010 Patient, Inv and Country data. Enrollment and Patient analytics
4 Psoriasis 6/2/2010 Patient, Inv and Country data. Enrollment and Patient analytics
5 Diabetes Mellitus 1 6/3/2010 Patient, Inv and Country data. wave 1, phase 3 trials - initial planning and enrollment
6 Diabetes Mellitus 2 6/26/2010 Patient, Inv and Country data. wave 1, phase 3 trials - initial planning and enrollment
7 Diabetes Mellitus Complete 6/28/2010 Patient, Inv and Country data. wave 1, phase 3 trials - initial planning and enrollment
8 Psoriasis Disease 6/28/2010 To be deleted.
9 Crohns Disease 6/28/2010 Patient data.
10 Head and Neck Cancer 7/23/2010 Inv and Country. Initial engagement with team evaluate feasibility of using ClinWeb for oncology.
11 SK&A US Rheumatologists 8/13/2010 2:06 PM Inv data only. Identify additional sites related to Rheumatoid Arthritis (SubjectID=11).
12 Colorectal Cancer 8/25/2010 Inv and Country data. Enrollment model
13 Hypogonadism 8/25/2010 Patient, Inv and Country data. Adhoc patient and enrollment questions.
14 Major Depressive Disorder 8/25/2010 Inv and Country data. Enrollment model.
16 Rheumatoid Arthritis September 8/31/2010 10:22 AM Inv and Country data. Reprediction of enrollment based on actual site start up. Identified additional sites.
17 Spondylosing Ankylosis 9/20/2010 To be deleted.
18 Psoriatic Arthritis 10/11/2010 To be deleted.
19 Cardiovascular 10/14/2010 Inv and Country data. Cardiomegie trials. Enrollment simulationof Routine Dat-outcomes.
20 Acute Coronary Syndrome 10/21/2010 Patient, Inv and Country data. Phase 2 retrospective enrollment.
21 Hypercholesterolemia 10/21/2010 Country data. Possible Phase 2 indication.
22 Phase 2 11/2/2010 1:42 PM Inv and Country data. Retrospective enrollment for phase 2 study
23 Coronary Heart Disease 1/18/201 Patient, Inv and Country data. Phase enrollment.
24 Alzheimer's Disease 2/25/2010 Patient, Inv and Country data. Demo & simulation of phase 2 study.
25 ACS and CHD 3/24/2010 Patient, Inv and Country data. Phase 3 enrollment.
26 Diabetic Kidney Disease 4/12/2011 Inv and country data. Support G-RFP on Lilly phase 3 study.
27 4/14/2011 Inv and Country data. Evaluate actual enrollment for
28 4/21/2011 Inv data. Evaluate actual enrollment for
29 Acute Coronary Syndrome 2011C 5/2/2011 Inv and Country data. After the ClinWeb demo using SubjectID=20 team provided input on this subject
30 Gastric Cancer 5/2/2011 To be deleted? Joe?
31 Rheumatoid Arthritis 2011105 5/17/2011 Patient, Inv and Country data. H2H enrollment predictions. This is a test data set.
32 Multiple Myeloma 5/23/2011 To be deleted? Joe?
Update Cycle Time and Staffing
Close

FIG. 4

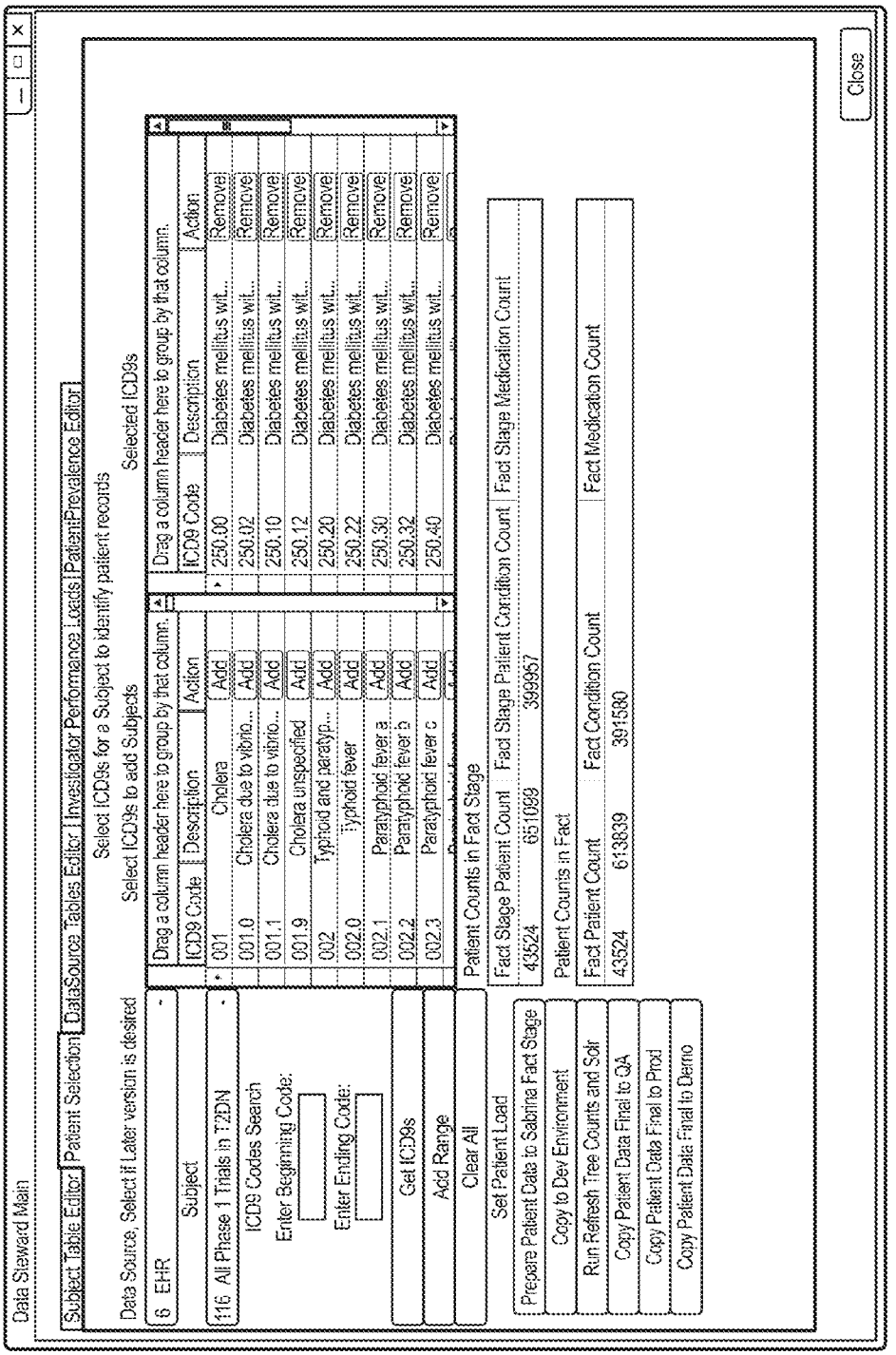

Data Steward Main
Subject Table Editor
Patient Selection
DataSource Tables Editor
Investigator Performance Loads
PatientPrevalence Editor
Select ICD9s for a Subject to identify patient records
Data Source, Select if Later version is desired
Select ICD9s to add Subjects
Selected ICD9s
6 EHR
Subject
116 All Phase 1 Trials in T2DN
ICD9 Codes Search
Enter Beginning Code:
Enter Ending Code:
Get ICD9s
Add Range
Clear All
Set Patient Load
Prepare Patient Data to Sabrina Fact Stage
Copy to Dev Environment
Run Refresh Tree Counts and Solr
Copy Patient Data Final to QA
Copy Patient Data Final to Prod
Copy Patient Data Final to Demo
Drag a column header here to group by that column.
ICD9 Code
Description
Action
001 Cholera Add
001.0 Cholera due to vibrio... Add
001.1 Cholera due to vibrio... Add
001.9 Cholera unspecified Add
002 Typhoid and paratyp... Add
002.0 Typhoid fever Add
002.1 Paratyphoid fever a Add
002.2 Paratyphoid fever b Add
002.3 Paratyphoid fever c Add
Drag a column header here to group by that column.
ICD9 Code
Description
Action
250.00 Diabetes mellitus wit... Remove
250.02 Diabetes mellitus wit... Remove
250.10 Diabetes mellitus wit... Remove
250.12 Diabetes mellitus wit... Remove
250.20 Diabetes mellitus wit... Remove
250.22 Diabetes mellitus wit... Remove
250.30 Diabetes mellitus wit... Remove
250.32 Diabetes mellitus wit... Remove
250.40 Diabetes mellitus wit... Remove
Patient Counts in Fact Stage
Fact Stage Patient Count
Fact Stage Patient Condition Count
Fact Stage Medication Count
43524
6510699
399957
Patient Counts in Fact
Fact Patient Count
Fact Condition Count
Fact Medication Count
43524
613839
391580
Close

FIG. 5

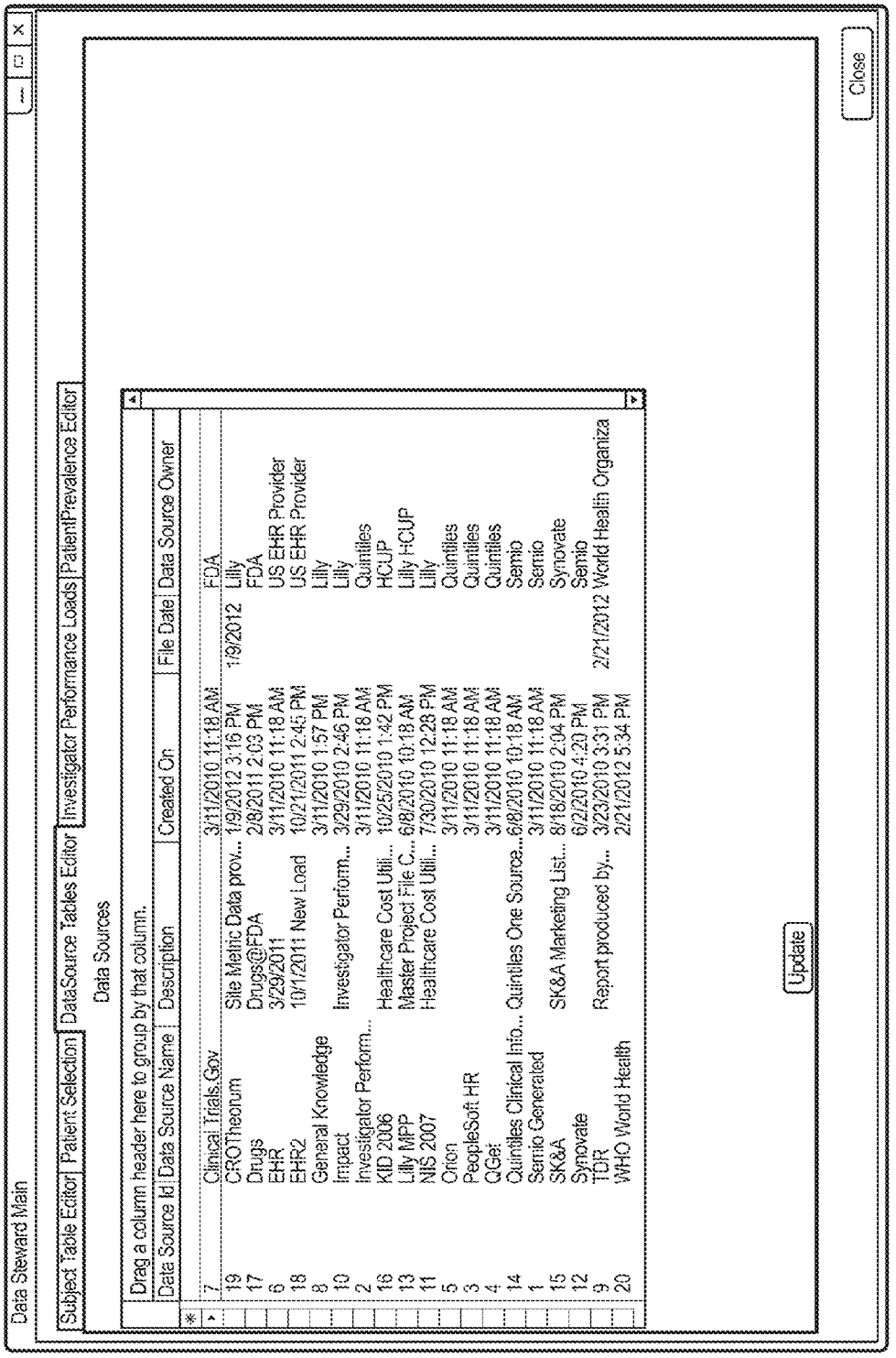

Data Steward Main
Subject Table Editor
Patient Selection
DataSource Tables Editor
Investigator Performance Loads
PatientPrevalence Editor
Data Sources
Drag a column header here to group by that column.
Data Source Id
Data Source Name
Description
Created On
File Date
Data Source Owner
7 Clinical Trials.Gov 3/11/2010 11:18 AM FDA
19 CROTheorum Site Metric Data prov... 1/9/2012 3:16 PM 1/9/2012 Lilly
17 Drugs Drugs@FDA 2/8/2011 2:03 PM FDA
6 EHR 3/29/2011 3/11/2010 11:18 AM US EHR Provider
18 EHR2 10/1/2011 New Load 10/21/2011 2:45 PM US EHR Provider
8 General Knowledge 3/11/2010 1:57 PM Lilly
10 Impact Investigator Perform... 3/29/2010 2:46 PM Lilly
2 Investigator Perform... 3/11/2010 11:18 AM Quintiles
16 KID 2006 Healthcare Cost Utili... 10/25/2010 1:42 PM HCUP
13 Lilly MPP Master Project File C... 6/8/2010 10:18 AM Lilly HCUP
11 NIS 2007 Healthcare Cost Utili... 7/30/2010 12:28 PM Lilly
5 Orion 3/11/2010 11:18 AM Quintiles
3 PeopleSoft HR 3/11/2010 11:18 AM Quintiles
4 QGet 3/11/2010 11:18 AM Quintiles
14 Quintiles Clinical Info... Quintiles One Source... 6/8/2010 10:18 AM Sernio
1 Sernio Generated 3/11/2010 11:18 AM Sernio
15 SK&A SK&A Marketing List... 8/18/2010 2:04 PM Synovate
12 Synovate 6/2/2010 4:20 PM Sernio
9 TDR Report produced by... 3/23/2010 3:31 PM 2/21/2012 World Health Organiza
20 WHO World Health 2/21/2012 5:34 PM
Update
Close

FIG. 6

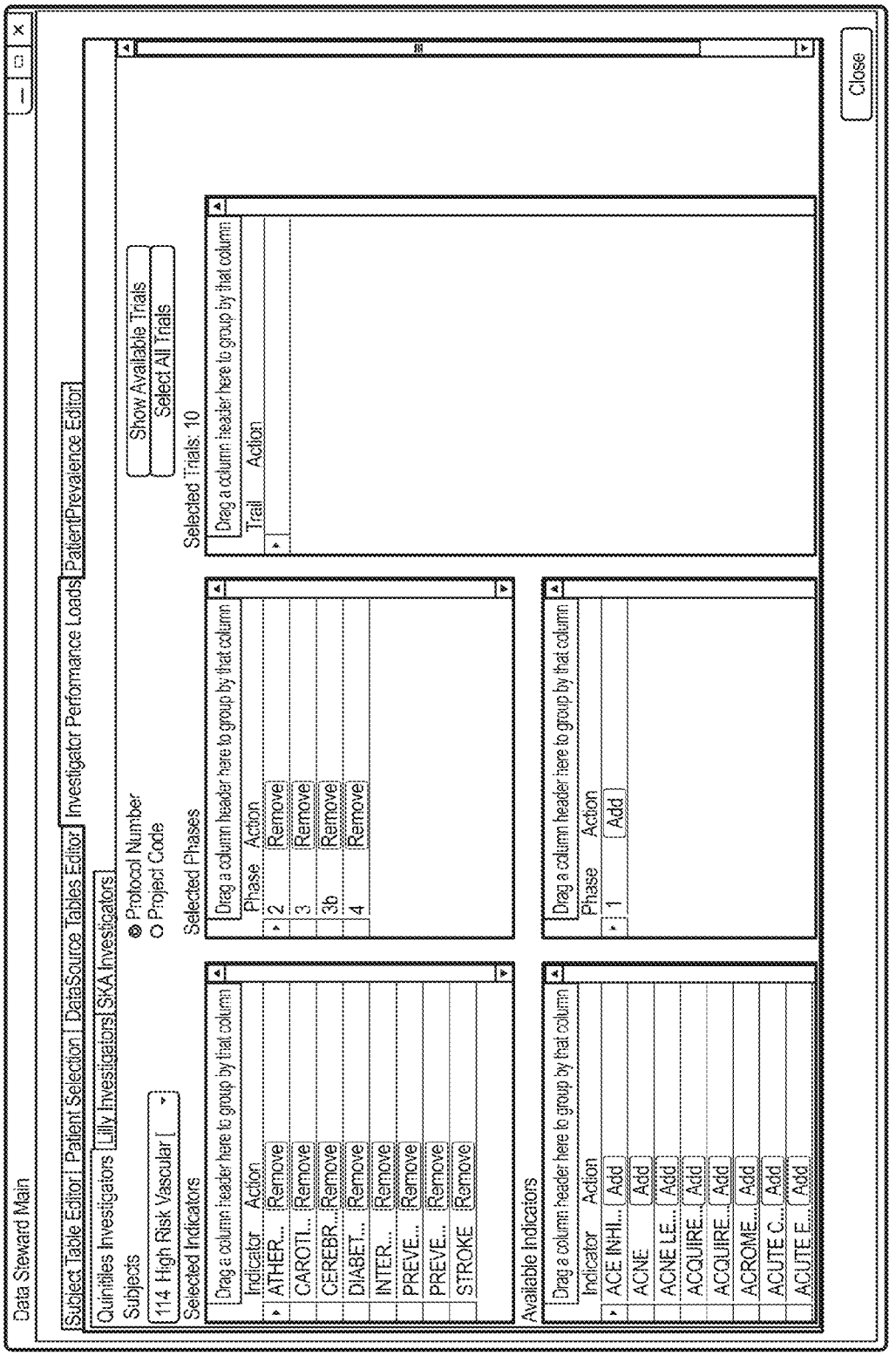

Data Steward Main
Subject Table Editor
Patient Selection
DataSource Tables Editor
Investigator Performance Loads
PatientPrevalence Editor
Quintiles Investigators
Lilly Investigators
SKA Investigators
Subjects
114 High Risk Vascular
Protocol Number
Project Code
Show Available Trials
Select All Trials
Selected Trials: 10
Selected Indicators
Selected Phases
Available Indicators
Drag a column header here to group by that column
Indicator
Action
Phase
Trail
ATHER...
CAROTI...
CEREBR...
DIABET...
INTER...
PREVE...
PREVE...
STROKE
ACE INHI...
ACNE
ACNE LE...
ACQUIRE...
ACQUIRE...
ACROME...
ACUTE C...
ACUTE E...
Remove
Add
2
3
3b
4
1
Close

FIG. 7

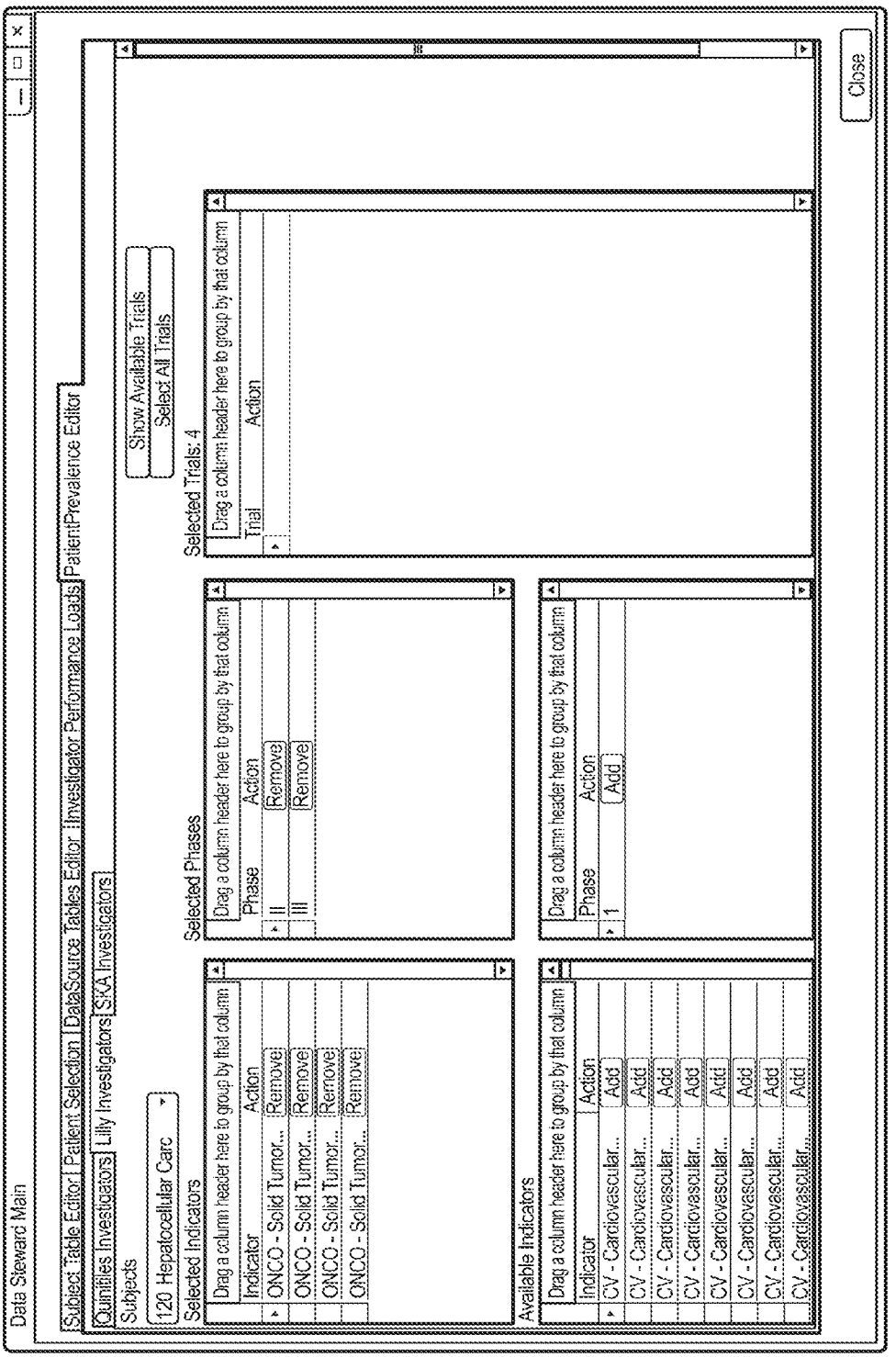

Data Steward Main
Subject Table Editor
Patient Selection
DataSource Tables Editor
Investigator Performance Loads
PatientPrevalence Editor
Quintiles Investigators
Lilly Investigators
SKA Investigators
Subjects
120 Hepatocellular Carc
Selected Indicators
Drag a column header here to group by that column
Indicator
Action
ONCO - Solid Tumor...
Remove
Available Indicators
CV - Cardiovascular...
Add
Selected Phases
Phase
II
III
1
Show Available Trials
Select All Trials
Selected Trials: 4
Trial
Close

FIG. 8

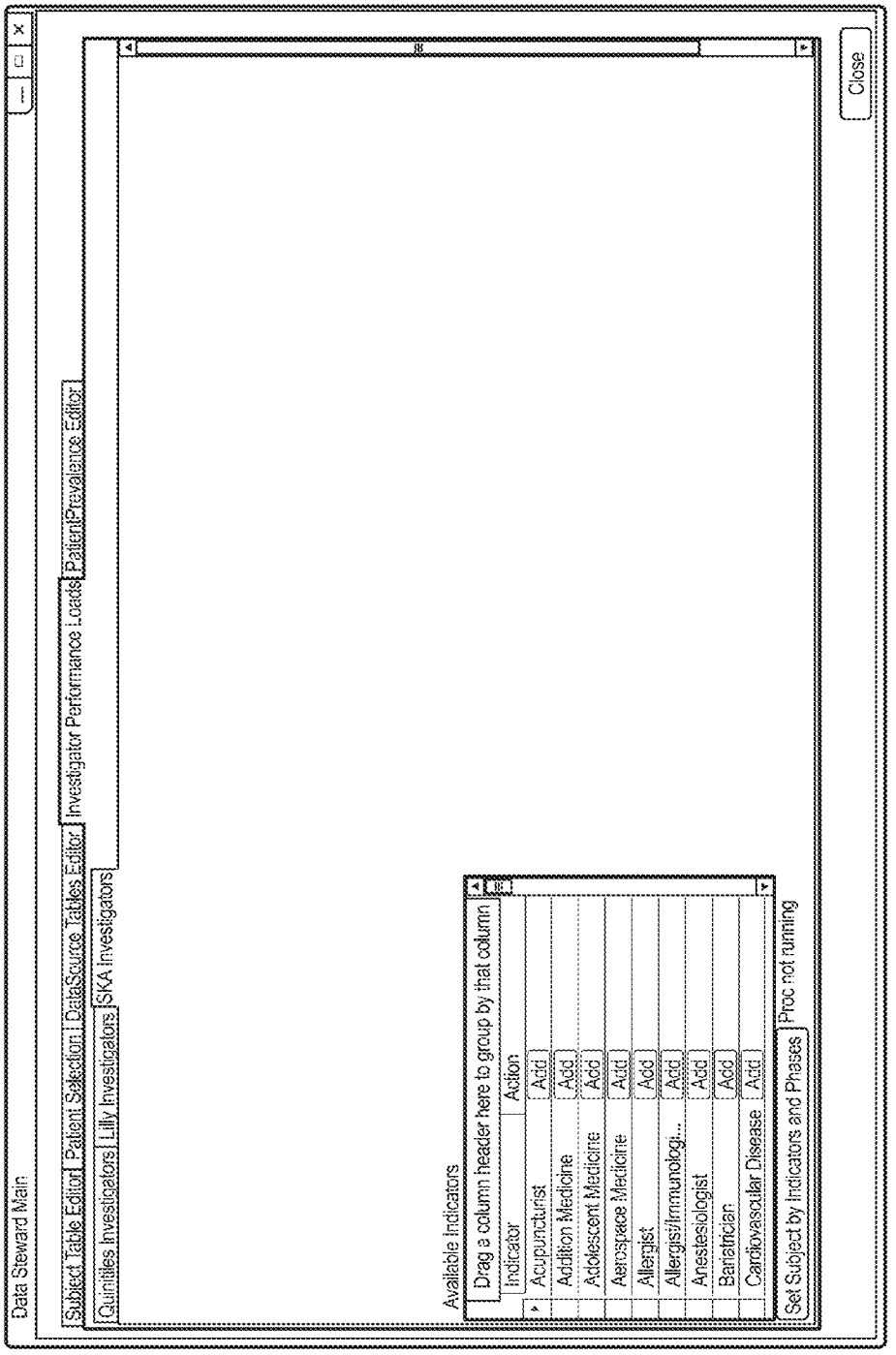

Data Steward Main
Subject Table Editor
Patient Selection
DataSource Tables Editor
Investigator Performance Loads
PatientPrevalence Editor
Quintiles Investigators
Lilly Investigators
SKA Investigators
Available Indicators
Drag a column header here to group by that column
Indicator
Action
Acupuncturist
Addiction Medicine
Adolescent Medicine
Aerospace Medicine
Allergist
Allergist/Immunologi...
Anesthesiologist
Bariatrician
Cardiovascular Disease
Add
Set Subject by Indicators and Phases
Proc not running
Close

FIG. 9

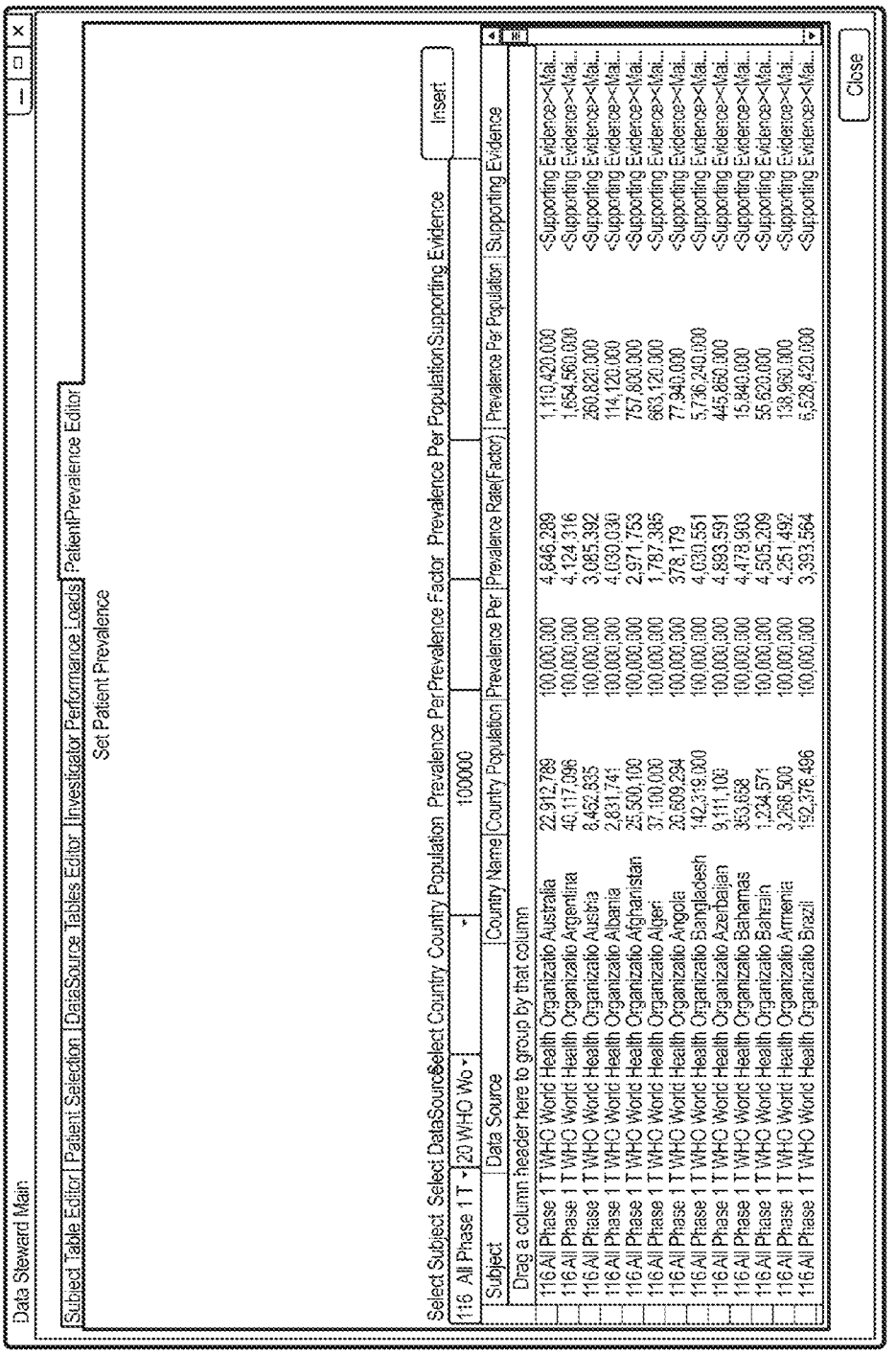

Data Steward Main
Subject Table Editor | Patient Selection | DataSource Tables Editor | Investigator Performance Loads | PatientPrevalence Editor
Set Patient Prevalence
Select Subject Select DataSource Select Country Country Population Prevalence Per Prevalence Factor Prevalence Per Population Supporting Evidence
Insert
1:16 All Phase 1 T
20 WHO Wo
100000
Subject
Data Source
Country Name
Country Population
Prevalence Per
Prevalence Rate(Factor)
Prevalence Per Population
Supporting Evidence
Drag a column header here to group by that column
Close

FIG. 10

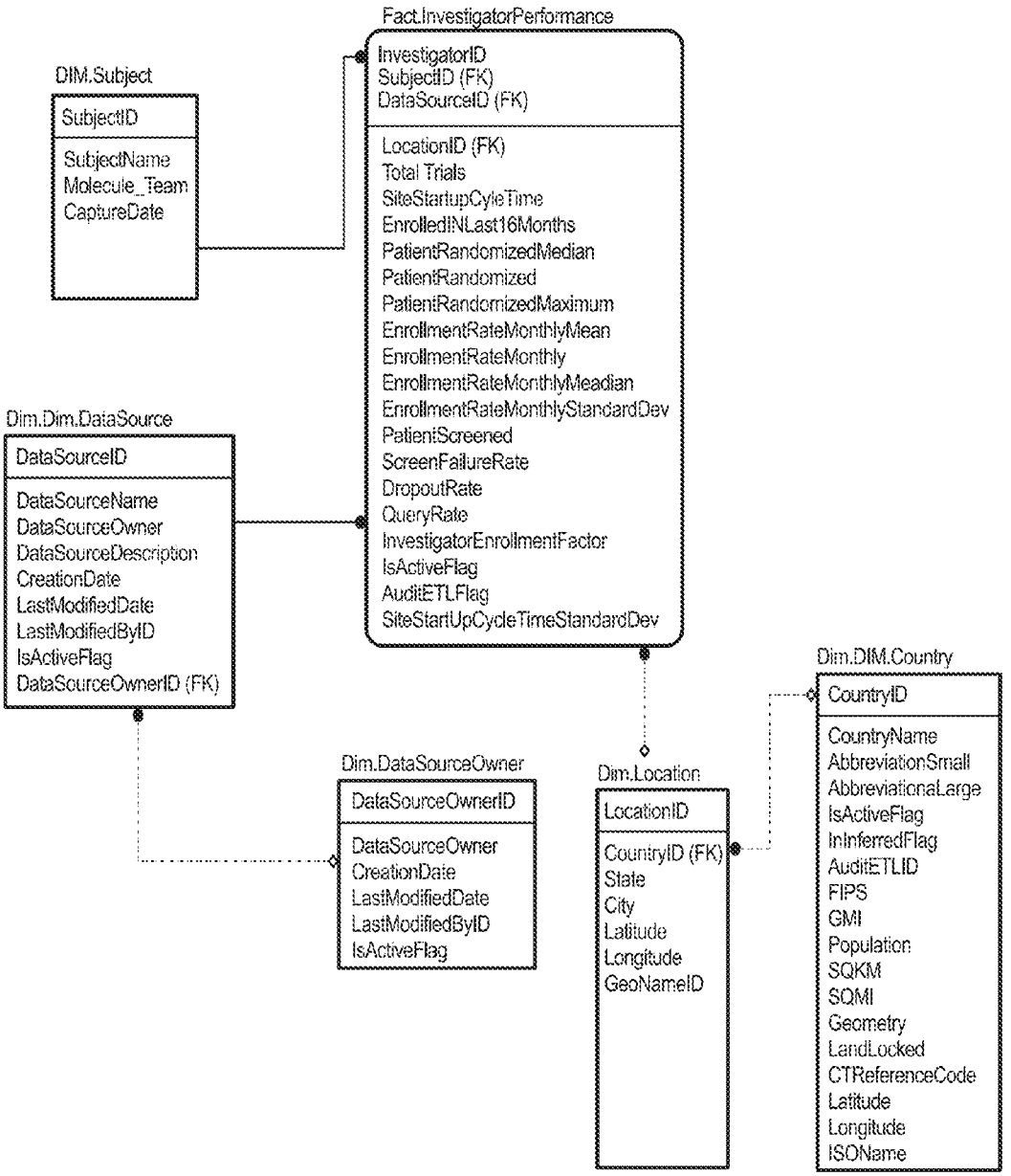
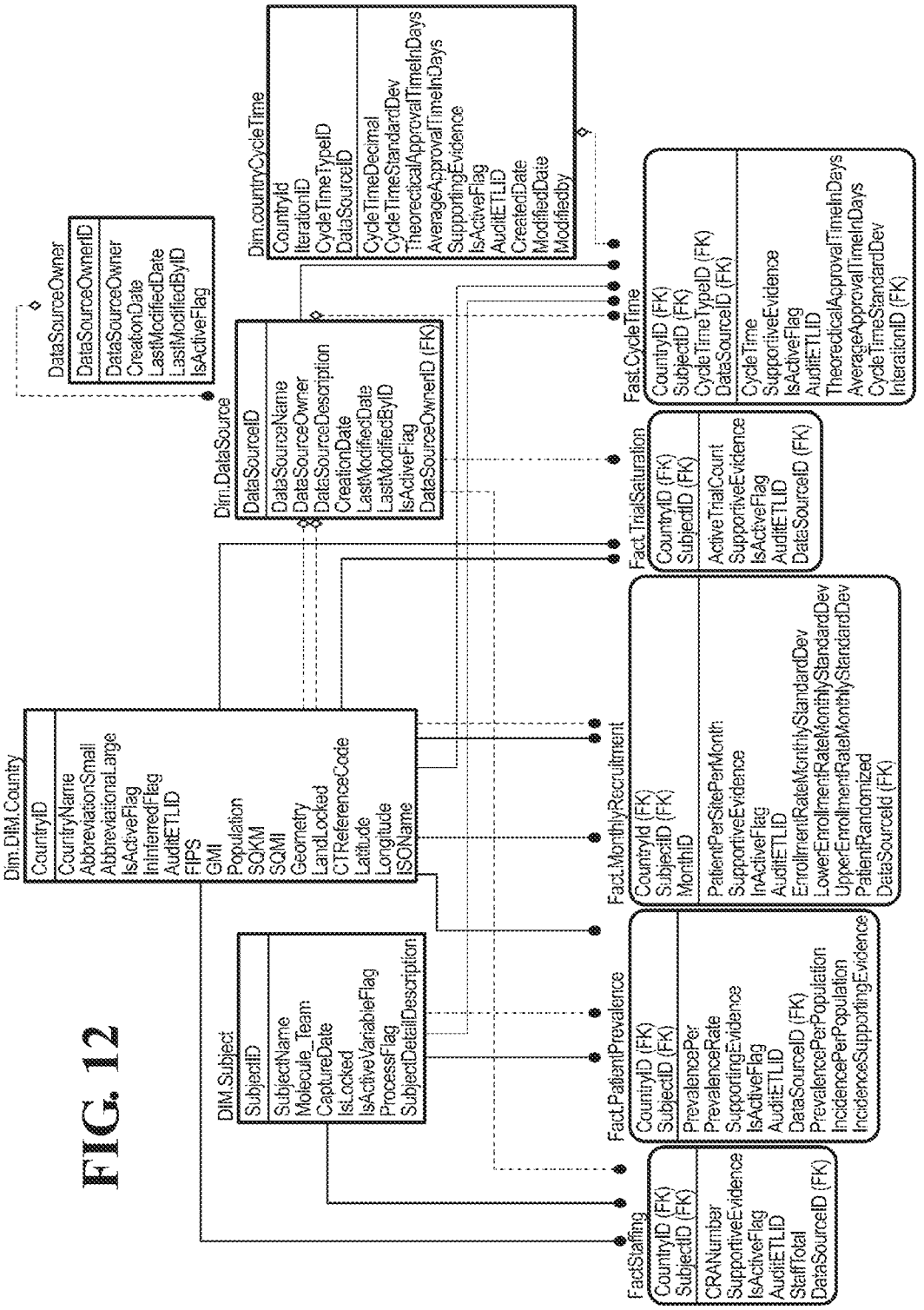

Fact.InvestigatorPerformance
InvestigatorID
SubjectID (FK)
DataSourceID (FK)
LocationID (FK)
Total Trials
SiteStartupCyleTime
EnrolledINLast16Months
PatientRandomizedMedian
PatientRandomized
PatientRandomizedMaximum
EnrollmentRateMonthlyMean
EnrollmentRateMonthly
EnrollmentRateMonthlyMeadian
EnrollmentRateMonthlyStandardDev
PatientScreened
ScreenFailureRate
DropoutRate
QueryRate
InvestigatorEnrollmentFactor
IsActiveFlag
AuditETLFlag
SiteStartUpCycleTimeStandardDev
DIM.Subject
SubjectID
SubjectName
Molecule_Team
CaptureDate
Dim.Dim.DataSource
DataSourceID
DataSourceName
DataSourceOwner
DataSourceDescription
CreationDate
LastModifiedDate
LastModifiedByID
IsActiveFlag
DataSourceOwnerID (FK)
Dim.DataSourceOwner
DataSourceOwnerID
DataSourceOwner
CreationDate
LastModifiedDate
LastModifiedByID
IsActiveFlag
Dim.Location
LocationID
CountryID (FK)
State
City
Latitude
Longitude
GeoNameID
Dim.DIM.Country
CountryID
CountryName
AbbreviationSmall
AbbreviationaLarge
IsActiveFlag
InInferredFlag
AuditETLID
FIPS
GMI
Population
SQKM
SQMI
Geometry
LandLocked
CTReferenceCode
Latitude
Longitude
ISOName

FIG. 11

FIG. 12


Dim.DIM.Country
CountryID
CountryName
AbbreviationSmall
AbbreviationLarge
IsActiveFlag
IsInferredFlag
AuditETLID
FIPS
GMI
Population
SQKM
SQMI
Geometry
LandLocked
CTReferenceCode
Latitude
Longitude
ISOName
DIM.Subject
SubjectID
SubjectName
Molecule_Team
CaptureDate
IsLocked
IsActiveVariableFlag
ProcessFlag
SubjectDetailDescription
FactStaffing
CountryID (FK)
SubjectID (FK)
CRANumber
SupportiveEvidence
IsActiveFlag
AuditETLID
StaffTotal
DataSourceID (FK)
Fact.PatientPrevalence
CountryID (FK)
SubjectID (FK)
PrevalenceRate
SupportingEvidence
IsActiveFlag
AuditETLID
DataSourceID (FK)
PrevalencePerPopulation
IncidencePerPopulation
IncidenceSupportingEvidence
Fact.MonthlyRecruitment
CountryId (FK)
SubjectID (FK)
MonthID
PatientPerSitePerMonth
SupportiveEvidence
IsActiveFlag
AuditETLID
EnrollmentRateMonthlyStandardDev
LowerEnrollmentRateMonthlyStandardDev
UpperEnrollmentRateMonthlyStandardDev
PatientRandomized
DataSourceId (FK)
Dim.DataSource
DataSourceID
DataSourceName
DataSourceOwner
DataSourceDescription
CreationDate
LastModifiedDate
LastModifiedByID
IsActiveFlag
DataSourceOwnerID (FK)
DataSourceOwner
DataSourceOwnerID
DataSourceOwner
CreationDate
LastModifiedDate
LastModifiedByID
IsActiveFlag
Dim.countryCycleTime
CountryId
IterationID
CycleTimeTypeID
DataSourceID
CycleTimeDecimal
CycleTimeStandardDev
TheoreticalApprovalTimeInDays
AverageApprovalTimeInDays
SupportingEvidence
IsActiveFlag
AuditETLID
CreatedDate
ModifiedDate
Modifiedby
Fact.TrialSaturation
CountryID (FK)
SubjectID (FK)
ActiveTrialCount
SupportiveEvidence
IsActiveFlag
AuditETLID
DataSourceID (FK)
Fact.CycleTime
CountryID (FK)
SubjectID (FK)
CycleTimeTypeID (FK)
DataSourceID (FK)
CycleTime
SupportiveEvidence
IsActiveFlag
AuditETLID
TheoreticalApprovalTimeInDays
AverageApprovalTimeInDays
CycleTimeStandardDev
IterationID (FK)

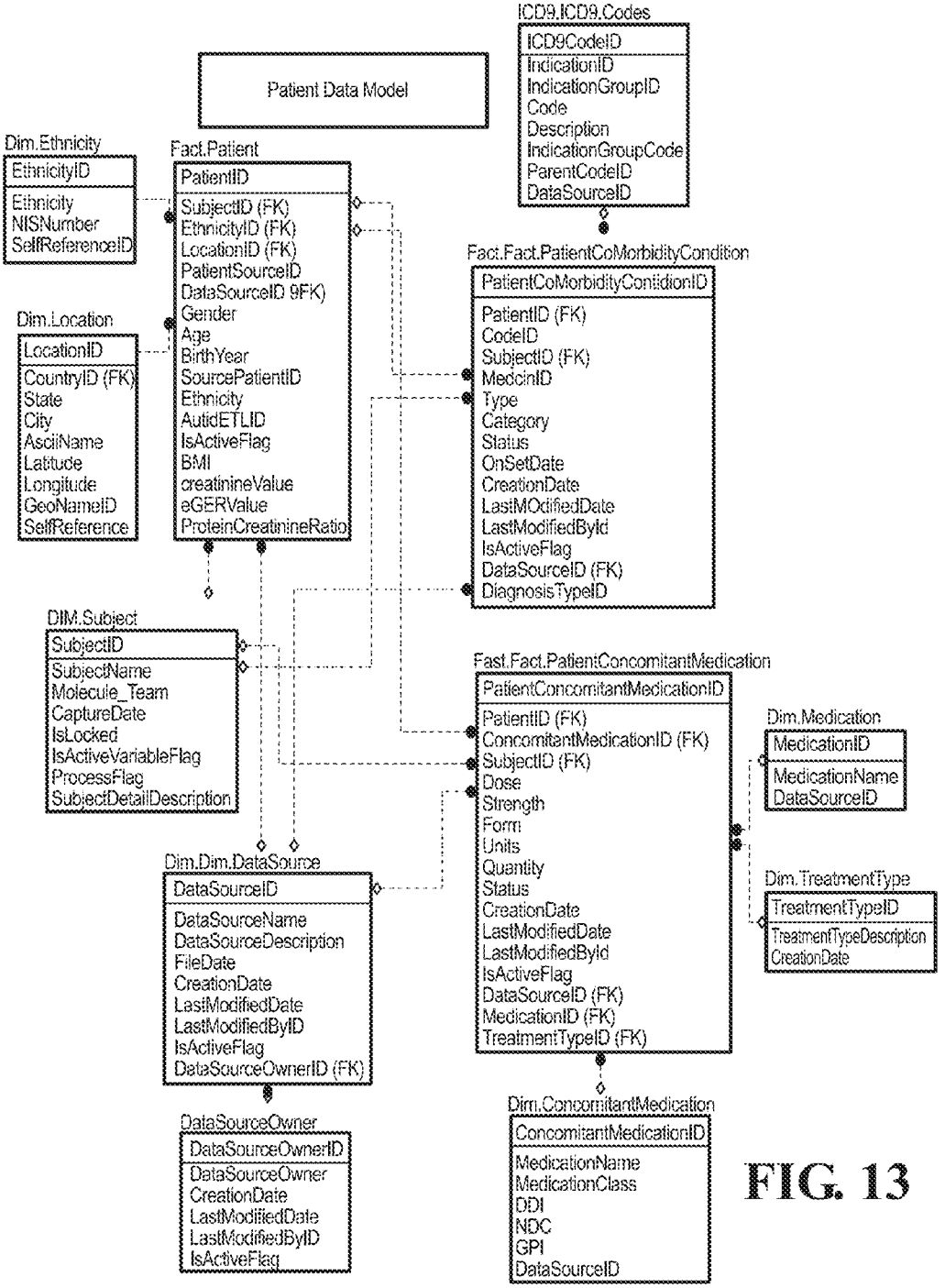

Patient Data Model
ICD9.ICD9.Codes
ICD9CodeID
IndicationID
IndicationGroupID
Code
Description
IndicationGroupCode
ParentCodeID
DataSourceID
Dim.Ethnicity
EthnicityID
Ethnicity
NISNumber
SelfReferenceID
Fact.Patient
PatientID
SubjectID (FK)
EthnicityID (FK)
LocationID (FK)
PatientSourceID
DataSourceID 9FK)
Gender
Age
BirthYear
SourcePatientID
Ethnicity
AutidETLID
IsActiveFlag
BMI
creatinineValue
eGERValue
ProteinCreatinineRatio
Fact.Fact.PatientCoMorbidityCondition
PatientCoMorbidityConiidionID
PatientID (FK)
CodeID
SubjectID (FK)
MedcinID
Type
Category
Status
OnSetDate
CreationDate
LastMOdifiedDate
LastModifiedById
IsActiveFlag
DataSourceID (FK)
DiagnosisTypeID
Dim.Location
LocationID
CountryID (FK)
State
City
AsciiName
Latitude
Longitude
GeoNameID
SelfReference
DIM.Subject
SubjectID
SubjectName
Molecule_Team
CaptureDate
IsLocked
IsActiveVariableFlag
ProcessFlag
SubjectDetailDescription
Fast.Fact.PatientConcomitantMedication
PatientConcomitantMedicationID
PatientID (FK)
ConcomitantMedicationID (FK)
SubjectID (FK)
Dose
Strength
Form
Units
Quantity
Status
CreationDate
LastModifiedDate
LastModifiedById
IsActiveFlag
DataSourceID (FK)
MedicationID (FK)
TreatmentTypeID (FK)
Dim.Medication
MedicationID
MedicationName
DataSourceID
Dim.TreatmentType
TreatmentTypeID
TreatmentTypeDescription
CreationDate
Dim.Dim.DataSource
DataSourceID
DataSourceName
DataSourceDescription
FileDate
CreationDate
LastModifiedDate
LastModifiedByID
IsActiveFlag
DataSourceOwnerID (FK)
DataSourceOwner
DataSourceOwnerID
DataSourceOwner
CreationDate
LastModifiedDate
LastModifiedByID
IsActiveFlag
Dim.ConcomitantMedication
ConcomitantMedicationID
MedicationName
MedicationClass
DDI
NDC
GPI
DataSourceID

FIG. 13

SYSTEMS AND METHODS FOR SUBJECT IDENTIFICATION (ID) MODELING

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to U.S. Provisional Application No. 61/663,292, filed on Jun. 22, 2012, entitled "Method and System to Manipulate Multiple Selections against a Population of Elements;" U.S. Provisional Application No. 61/663,057, filed on Jun. 22, 2012, entitled "Systems and Methods For Predictive Analytics For Site Initiation and Patient Enrollment;" U.S. Provisional Application No. 61/663,299, filed on Jun. 22, 2012, entitled "Methods and Systems for Predictive Clinical Planning and Design and integrated Execution Services;" U.S. Provisional Application No. 61/663,398, filed on Jun. 22, 2012, entitled "Systems and Methods for Subject Identification (ID) Modeling;" U.S. Provisional Application No. 61/663,219, filed Jun. 22, 2012, entitled "Systems and Methods for Analytics on Viable Patient Populations;" U.S. Provisional Application No. 61/663,357, filed Jun. 22, 2012; entitled "Methods and Systems for a Clinical Trial Development Platform;" U.S. Provisional Application No. 61/663,216, filed Jun. 22, 2012; entitled "Systems and Methods for Data Visualization." The entirety of all of which is hereby incorporated by reference herein.

FIELD OF THE INVENTION

[0002] The present invention relates generally to systems and methods for the creation and analysis of data associated with clinical trials. The present invention relates more specifically to systems and methods for subject identification (ID) modeling.

BACKGROUND

[0003] Clinical trials for molecules that may become pharmaceutical products often last for years. The core cost of the that is affected primarily by the length of the trial. And a delay of even a single day can cost hundreds or thousands and even millions of dollars.

[0004] Data associated with clinical trials is often associated with various data sources and may be highly diverse. Highly diverse data from numerous data sources is often difficult to organize, assemble, and analyze. Therefore, systems and methods for the collation of highly diverse data into usable data would be advantageous. Furthermore, systems and methods for dynamic analysis to support better understanding of the impacts of decisions against clinical trials would be advantageous.

SUMMARY

[0005] Embodiments of the present invention provide systems and methods for subject identification (ID) modeling. In one embodiment, raw data is processed to an application using a tool that enables a user to build subject identifications dynamically. In some embodiments, such subject IDs can be created by a user without technical expertise.

[0006] In one embodiment, a subject identification can be created dynamically. For example, a user can interact with a user interface to create a subject by entering or selecting a subject name and a molecule team. In one embodiment, a unique subject identification ID is automatically created or assigned once a subject name and a molecule team have been entered or selected.

[0007] A subject identification may be associated with information contained in other tables and/or databases. For example, in one embodiment, subject identifications may be associated with information contained in one or more core domains such as a patient domain, a country domain, and/or an investigator domain. In some embodiments, one or more domains are generically designed such that data sources that are unknown at the time the domain(s) are created can be managed. In this way, using a generic structure that supports a domain, data sources can be added and/or updated as additional information and/or data sources become available. Exemplary models that depict generic structures which support such domains are disclosed herein and variations are within the scope of this disclosure.

[0008] Using various graphical user interfaces, a user can dynamically associate patient criteria, country criteria, investigator criteria, and/or other information with subject identifications. In some embodiments, as the user interacts with the graphical user interfaces, associations between subject identifications and data in other tables and/or databases is dynamically updated in real-time or substantially real time. For example, when a user selects an indicator to be associated with a particular subject identification, the association may be created. As another example, when a user selects various indicators to be associated with a subject identification and then clicks an update button on the graphical user interface, the selected indicators may be dynamically associated with the subject identification.

[0009] Information associated with a particular subject identification may be frozen at a particular date and/or time. For example, a subject identification may be associated with a specified capture date. In this embodiment, information contained in the various domains may be filtered such that only information contained in the domain on or before the capture date is available for the subject identification. In this way, information for a data model may be updated as additional information for the data model becomes available but the information available to a particular subject identification can be limited to a static point in time.

[0010] These embodiments are mentioned not to limit or define the invention, but to provide an example of an embodiment of the invention to aid understanding thereof. Embodiments are discussed in the Detailed Description, and further description of the invention is provided there. Advantages offered by the various embodiments of the present invention may be further understood by examining this specification.

BRIEF DESCRIPTION OF THE FIGURES

[0011] These and other features, aspects, and advantages of the present invention are better understood when the following Detailed Description is read with reference to the accompanying drawings, wherein:

[0012] FIG. **1** is a block diagram illustrating an exemplary environment for implementation of one embodiment of the present invention;

[0013] FIG. **2** is a flowchart illustrating a method for dynamically creating and/or updating information associated with subject identifications according to one embodiment of the present invention;

[0014] FIG. **3** is a partial entity-relationship diagram illustrating how a subject ID is linked to complex and/or varied data sets from multiple sources according to one embodiment of the present invention;

[0015] FIG. **4** is a screen shot of a subject table editor according to one embodiment of the present invention;

[0016] FIG. **5** is a screen shot of patient selection according to one embodiment of the present invention;

[0017] FIG. **6** is a screen shot of a data source table editor according to one embodiment of the present invention;

[0018] FIG. **7** is a screen shot of investigators for investigator performance loads according to embodiments of the present invention;

[0019] FIG. **8** is a screen shot of investigators for investigator performance loads according to embodiments of the present invention;

[0020] FIG. **9** is a screen shot of investigators for investigator performance loads according to embodiments of the present invention;

[0021] FIG. **10** is a screen shot of a patient prevalence editor according to one embodiment of the present invention

[0022] FIG. **11** is an exemplary investigator data model according to one embodiment of the present invention;

[0023] FIG. **12** is an exemplary country data model according to one embodiment of the present invention; and

[0024] FIG. **13** is an exemplary patient data model according to one embodiment of the present invention.

DETAILED DESCRIPTION

[0025] Embodiments of the present invention provide systems and methods for subject identification (ID) modeling.

Illustrative Embodiment of the Present Invention

[0026] One illustrative embodiment of the present invention comprises an application for creating and/or updating subject identification (ID) models for clinical trials. The embodiment allows a user to access an application that presents a variety of clinical trial-related parameters for various patients, countries, and/or investigators. These parameters may include, for example, the population of a country, the regulatory environment, and/or the level of risk associated with conducting a trial in a particular country.

[0027] Using various graphical user interfaces associated with the application, a user can create subject identifications and/or select parameters related to patients, countries, and/or investigators for subject identifications. For example, in one embodiment, a user can select patients associated with one or more ICD9 codes for a particular subject identification. As another example, a user can add or remove various indicators—such as Ace Inhibitors, Acne, etc.—and/or phases and/or trials.

[0028] As the user interacts with the graphical user interface associations with data contained in various databases are added, removed, or updated. In some embodiments, such associations may be added, removed, or updated dynamically without requiring technical expertise regarding the underlying data structures. For example, a user can select one or more ICD9 codes to associate with a particular subject identification. In one embodiment. When a user selects an ICD9 code, associations between the selected ICD9 code, patients corresponding to the ICD9 code, and/or the subject identification are created. Similarly, when a user deselects an ICD9 code, associations between subject identifications and information in other tables and/or databases may be updated or removed.

[0029] In the illustrative embodiment, for various investigators, the user is able to specify investigator-specific parameters, such as indicators, phases, trials and/or other relevant parameters. The process is iterative; the user is able to change the parameters for patients, countries and/or investigators to determine the most appropriate sites to utilize for a clinical trial. As the user changes these parameters, information and/or associations corresponding to subject identifications and/or data structures may be dynamically added, removed, or updated. The results of the user's selections can then be used as part of a larger clinical trial analysis application.

[0030] This illustrative embodiment neither limits nor defines the invention. Rather, the illustrative embodiment is meant to provide an example of how the present invention may be implemented.

Illustrative Environment

[0031] Referring now to the drawings, in which like numerals indicate like elements throughout the several figures, FIG. **1** is a block diagram illustrating an exemplary environment for implementation of one embodiment of the present invention. The embodiment shown in FIG. **1** includes a client **100** that allows a user to interface with an application server **200**, web server **300**, and/or database **400** via a network **500**.

[0032] The client **100** may be, for example, a personal computer (PC), such as a laptop or desktop computer, which includes a processor and a computer-readable media. The client **100** also includes user input devices, such as a keyboard and mouse or touch screen, and one or more output devices, such as a display. In some embodiments of the invention, the user of client **100** accesses an application or applications specific to one embodiment of the invention. In other embodiments, the user accesses a standard application, such as a web browser on client **100**, to access applications running on a server such as application server **200**, web server **300**, or database **400**. For example, in one embodiment, in the memory of client **100** are stored applications including a design studio application for planning and designing clinical trials. The client **100** may also be referred to as a terminal in some embodiments of the present invention.

[0033] Such applications may be resident in any suitable computer-readable medium and executable on any suitable processor. Such processors may comprise, for example, a microprocessor, an ASIC, a state machine, or other processor, and can be any of a number of computer processors, such as processors from Intel Corporation, Advanced Micro Devices Incorporated, and Motorola Corporation. The computer-readable media stores instructions that, when executed by the processor, cause the processor to perform the steps described herein.

[0034] The client **100** provides a software layer, which is the interface through which the user interacts with the system by receiving and displaying data to and from the user. In one embodiment, the software layer is implemented in the programming language C# (also referred to as C Sharp). In other embodiments, the software layer can be implemented in other languages such as Java or C++. The software layer may be graphical in nature, using visual representations of data to communicate said data to one or more users. The visual representations of data may also be used to receive additional data from one or more users. In one embodiment, the visual

representation appears as a spider-like layout of nodes and connectors extending from each node to a central node.

[0035] Embodiments of computer-readable media comprise, but are not limited to, an electronic, optical, magnetic, or other storage device, transmission device, or other device that comprises some type of storage and that is capable of providing a processor with computer-readable instructions. Other examples of suitable media comprise, but are not limited to, a floppy disk, CD-ROM, DVD, magnetic disk, memory chip, ROM, RAM, PROM, EPROM, EEPROM, an ASIC, a configured processor, all optical media, all magnetic tape or other magnetic media, or any other medium from which a computer processor can read instructions. Also, various other forms of computer-readable media may be embedded in devices that may transmit or carry instructions to a computer, including a router, private or public network, or other transmission device or channel, both wired and wireless. The instructions may comprise code from any suitable computer-programming language, including, for example, C, C#, Visual Basic, Java, Python, Perl, and JavaScript.

[0036] The application server **200** also comprises a processor and a memory. The application server may execute business logic or other shared processes. The application server may be, for example, a Microsoft Windows Server operating in a NET framework, an IBM Weblogic server, or a Java Enterprise Edition (J2E) server. While the application server **200** is shown as a single server, the application server **200**, and the other servers **300**, **400** shown may be combined or may include multiple servers operating together to perform various processes. In such embodiments, techniques such as clustering or high availability clustering may be used. Benefits to architectures such as these include redundancy and performance, among others.

[0037] In the embodiment shown in FIG. **1**, the application server **200** is in communication with a web server **300** via a network connection **250**. The web server **300** also comprises a processor and a memory. In the memory are stored applications including web server software. Examples of web server software include Microsoft Internet Information Services (IIS), Apache Web Server, and Sun Java System Web Server from Oracle, among others.

[0038] In the embodiment shown in FIG. **1**, the web server **300** is in communication with a database **400** via a network connection **350** and a network connection **450**. The web server **300** provides a web service layer that, together or separate from application server **200**, acts as middleware between a database **400** and the software layer, represented by the client **100**. The web server **300** communicates with the database **400** to send and retrieve data to and from the database **400**.

[0039] The network **500** may be any of a number of public or private networks, including, for example, the Internet, a local area network ("LAN"), or a wide area network ("WAN"). The network connections **150**, **250**, **350**, and **450** may be wired or wireless networks and may use any known protocol or standard, including TCP/IP, UDP, multicast, 802.11b, 802.11g, 802.11n, or any other known protocol or standard. Further, the network **100** may represent a single network or different networks. As would be clear to one of skill in the art, the client **100**, servers **200**, **300**, and database **400** may be in communication with each other over the network or directly with one another.

[0040] The database **400** may be one or a plurality of databases that store electronically encoded information comprising the data required to plan, design, and execute a clinical trial. In one embodiment, the data comprises one or more design elements conesponding to the various elements related to one or more clinical trials. The database **400** may be implemented as any known database, including an SQL database or an object database. Further, the database software may be any known database software, such as Microsoft SQL Server, Oracle Database, MySQL, Sybase, or others.

Illustrative Process for Adding/Updating Information for Subject Identifications

[0041] FIG. **2** is a flowchart illustrating a method for dynamically creating and/or updating information associated with subject identifications according to one embodiment of the present invention.

[0042] The method **200** shown in FIG. **2** begins when a subject is created or selected **210**. For example, an application may comprise a user interface for creating or selecting a subject. In one embodiment, a subject can be created by entering or selecting a subject name and/or a molecule team. In such an embodiment, a unique subject identification corresponding to the entered or selected subject name and/or molecule team may be dynamically created. For example, in one embodiment, when a cursor moves off a particular row, data associated with the row may be added or updated in a database and/or table containing information for subjects. In one embodiment, a message is shown in the application that indicates whether the database and/or table were successfully modified.

[0043] After creating or selecting a subject **210**, the method **200** proceeds to block **220**. In block **220**, patient parameters are received. For example, codes corresponding to particular diseases and/or illnesses, such as ICD9 or ICD10 codes, may be received through a graphical user interface of the application. Based on the selected patient parameters, a number of patients in one or more databases that meet the selected criteria may be displayed in the application. In some embodiments, information regarding patients and/or patient parameters may be stored in multiple databases and/or tables. In such an embodiment, the application may dynamically create or update associations between the various databases and/or tables containing information corresponding to patients and selected subject identification(s). For example, patient information may be stored in one or more patient databases corresponding to one or more generic patient data models. In such an embodiment, associations between the patient database(s) and subject may be dynamically created, modified, or removed based on selected or deselected patient parameters.

[0044] After patient parameters are received **220**, the method **200** proceeds to block **230**. In block **230**, investigator parameters are received. For example, various indicators, phases, and/or trials may be selected or removed for one or more subject identifications. In some embodiments, information regarding investigators and/or investigators parameters may be stored in multiple databases and/or tables. In such an embodiment, the application may dynamically create or update associations between the various databases and/or tables containing information corresponding to investigators and selected subject identification(s). For example, investigator information may be stored in one or more investigator databases corresponding to one or more generic investigator data models. In such an embodiment, associations between

the investigator database(s) and subjects may be dynamically created, modified, or removed based on selected or deselected investigator parameters.

[0045] After investigator parameters are received **230**, the method **200** proceeds to block **240**. In block **240**, country parameters are received. In some embodiments, information regarding countries and/or country parameters may be stored in multiple databases and/or tables. In such an embodiment, the application may dynamically create or update associations between the various databases and/or tables containing information corresponding to countries and selected subject identification(s). For example, country information may be stored in one or more investigator databases corresponding to one or more generic country data models. In such an embodiment, associations between the country database(s) and subjects may be dynamically created, modified, or removed based on selected or deselected country parameters.

[0046] After country parameters are received **240**, the method **200** proceeds to block **250**. In block **250**, a capture date is received. Based on the capture date for a particular subject, the application may update the subject identification such that only information associated with the databases on or before the capture date can be included in any data analysis associated with the subject. For example, one or more databases associated with a subject may periodically be updated as new information becomes available. In this embodiment, even if the database is updated, the information available for a selected subject may be limited to information associated with the database on or before the capture date associated with the subject. In this way, a static snapshot of data for a particular subject can be maintained. While allowing databases to continue to receive additional data as it becomes available. In some embodiments, the additional information may be used by other subjects that either do not have a specified capture date or that have a capture date that is after the information is received. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

Illustrative Associations for Subject ID and Data Sets

[0047] FIG. **3** is a partial entity-relationship diagram illustrating how a subject ID is linked to complex and/or varied data sets from multiple sources according to one embodiment of the present invention. In the embodiment shown in FIG. **3**, subject identifications (IDs), subject names, molecule teams, and/or capture dates are stored in a subject table. In this embodiment, a unique subject ID may be selected or assigned for each subject. In one embodiment, one or more of the subject name, molecule team, and/or capture date may be optional. In other embodiments, one or more of the subject name, molecule team, and/or capture date may be required. In still other embodiments, additional information for a subject ID may be optional or required. Referring back to FIG. **3**, numerous tables contain data that corresponds to subject IDs in the subject table. For example, in FIG. **3**, a subject ID in the Subject (Dim) table may correspond with a subject IL) in the Staffing (Fact) table. In this embodiment, the subject ID in the Staffing (Fact) table corresponds with a country identification. In other embodiments, information such as staffing, cycle times, monthly recruitment, trial saturation, patient prevalence, patient information, patient morbidity, patient concomitant information, investigator performance information, and/or investigator trial information—may be associated with a subject identification (ID) in a subject table. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

Data Steward Application and Screen Shots

[0048] Data Steward is a software tool that can be used to directly modify databases according to embodiments of the present invention. Appendix A, which is hereby incorporated by reference in its entirety, comprises a User Guide for a Data Steward according to embodiments of the present invention.

[0049] FIG. **4** is a screen shot of a subject table editor according to one embodiment of the present invention. In the subject table editor interface of the Data Steward application, a user can sort information displayed in the subject table editor interface based on subject identifications, subject names, molecule teams, capture dates, whether the subject ID is located, whether a subject ID has been processed, descriptions, and/or other information associated with a subject ID. In the embodiment shown in FIG. **4**, a user can lock or unlock a particular subject ID by selecting or deselecting a checkbox con to a particular subject ID. Likewise, in FIG. **4**, a user can selected whether or not a subject is processed by selecting or deselecting a checkbox corresponding to a particular subject ID. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

[0050] FIG. **5** is a screen shot of patient selection according to one embodiment of the present invention. In the patient selection interface of the Data Steward application, a user can select ICD codes for a subject such that matching patient records can be identified. In the embodiment shown in FIG. **5**, a user can select a subject identification and a data source for the subject identification. In this embodiment, a user can add and/or remove ICD9 codes. In other embodiments, other information which identifies a disease, illness, or other intbrmation usable to filter patient records may be selected. In the embodiment shown in FIG. **5**, information regarding patient counts, such as patient counts in fact stage and/or patient counts in facts, may be displayed based at least in part on the selected information. For example, in FIG. **5**, an Electronic Health Records (EHR) data source has been selected as well as various ICD9 codes. In this embodiment, a patient count, condition count, and medication count are displayed based at least in part on the selected data source and the selected ICD9 codes. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

[0051] FIG. **6** is a screen shot of a data source table editor according to one embodiment of the present invention. In the embodiment shown in FIG. **7**, information corresponding to various data sources is displayed. For example, a data source in the DataSource Tables Editor can have a corresponding data source identification, data source name, description, create date, file date, and/or data source owner. In other embodiments, additional or less information for one or more data sources is shown in the DataSource Tables Editor user interface. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

[0052] FIGS. **7**, **8**, and **9** are screen shots of investigators for investigator performance loads according to embodiments of the present invention. These screenshots illustrate performance load criteria that may be selected or removed for a particular subject identification according to embodiments. For example, various indicators, phases, and/or trials may be selected or removed for one or more subject identifications.

Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

[0053] FIG. **10** is a screen shot of a patient prevalence editor according to one embodiment of the present invention. In the embodiment shown in FIG. **10**, a user can filter information from various databases based on one or more of the following: subjects, data sources, countries, country population, prevalence, prevalence factor, prevalence per population, and/or supporting evidence. Data displayed in the graphical user interface of the Data Steward application may be sorted by subject, data source, country name, country population, prevalence, prevalence rate, prevalence per population, and/or supporting evidence. In the embodiment shown in FIG. **10**, a column header can be dragged to a particular location of the user interface to group information by that column. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

Exemplary Investigator Data Model

[0054] FIG. **11** is an exemplary investigator data model according to one embodiment of the present invention. In the embodiment shown in FIG. **11**, a FactInvestigatorPerformance table contains information associated with the patient randomization, enrollment rates, and failure rates as well as information associated with investigator identifications, subject identifications, data source identifications, and location identifications which correspond with additional information contained in other tables and/or databases. For example, a particular subject identification (SubjectID) in the FactInvestigatorPerformance table may correspond with a SubjectID in the DIM.Subject table.

[0055] By using the SubjectID in the FactinvestigatorPerformance table to query the SubjectID in the DIM.Subject table, additional information such as the subject name, molecule team, and/or capture can be determined for the SuhiectID contained in the FactinvestigatorPerformance table. In some embodiments, information contained in various tables and/or databases may be linked in a chain. For example, a DataSourceID in the FactinvestigatorPerformance table correspond with a DataSourceID in the Dim.Dim.DataSource table. The Dim.Dim.DataSource table, in turn, may contain a DataSourceOwnerID fro a particular DataSourceID which corresponds with a DataSourceOwnerID in the Dim.DataSourceOwner table. Thus, by querying the various tables and/or databases, a DataSourceID in the FactinvestigatorPerformance table can be used to determine information such as the LastModifiedDate in the Dim. DataSourceOwner table for the DataSource associated with the DataSourceID. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

[0056] Below is a description of the various tables of an investigator data model according to one embodiment of the present invention:

List of Tables in Investigator Data Model

Name	Comment
DataSourceOwner	
DIM.Country	Conformed country dimension consisting of ISO-3166 standards and minor manual updates to streamline country presentation data for Semio.

-continued

List of Tables in Investigator Data Model

Name	Comment
Dim.DataSource	
Dim.Location	Location dimension table contains location information about country, state, city, latitude, and longitude.
DIM.Subject	Subject table is the dimensional table used to snapshot fact data by molecule team, search phrase, and date.
Fact.InvestigatorPerformance	Fact information about Investigator performance.

Column(s) of "DataSourceOwner" Table

Name	Datatype	Comment	Is PK	Is FK
DataSourceOwnerID	int		Yes	No
DataSourceOwner	nvarchar(50)		No	No
CreationDate	datetime		No	No
LastModifiedDate	datetime		No	No
LastModifiedByID	varbinary(85)		No	No
IsActiveFlag	bit		No	No

Column(s) of "DIM.Country" Table

Name	Datatype	Comment	Is PK	Is FK
CountryID	integer	Country Identification Number. The primary key	Yes	No
CountryName	varchar( )	Country Description	No	No
AbbreviationSmall	nchar(2)	Two letter abbreviation name of the country	No	No
AbbreviationLarge	nchar(3)	Three letters name of the country	No	No
IsActiveFlag	char(18)	bit indicator for the validity of the record	No	No
IsInferredFlag	char(18)		No	No
AuditETLID	char(18)	Reference to Audit. ExecutionLog key for auditing	No	No
FIPS	nvarchar(255)	Two lettres code for the country	No	No
GMI	nvarchar(255)	Three lettres code for the country	No	No
Population	bigint	Population of the country	No	No
SQKM	float	The total square kilometer of a country	No	No
SQMI	float	The total miles of a country	No	No
Geometery	geometry	Geomatrical information of the country	No	No
LandLocked	char(1)	The country has a landlocked or not?	No	No
CTReferenceCode	varchar(10)	Clinical Trial reference code	No	No
Latitude	decimal(19, 12)	Latitude information of the country	No	No
Longitude	decimal(19, 12)	Longitude information of the country	No	No
ISOName	nvarchar(255)		No	No

Column(s) of "Dim.DataSource" Table

Name	Datatype	Comment	Is PK	Is FK
DataSourceID	int	Data Source Unique Identity Number	Yes	No
DataSourceName	nvarchar(50)	Name of the Data Source Provider	No	No
DataSourceOwner	Nvarchar(100)	Name of the Data Source Owner	No	No
DataSourceDe-scription	nvarchar(255)	Description of the Data Source	No	No
CreationDate	datetime	The date on which the data is being inserted	No	No
LastModifiedDate	datetime	Last modified date	No	No
LastModifiedByID	varbinary(85)	Last modified by Identity number	No	No
IsActiveFlag	bit	Bit indicator for the validity of the record	No	No
DataSourceOwnerID	int		No	Yes

Column(s) of "Dim.Location" Table

Name	Datatype	Comment	Is PK	Is FK
LocationID	int	Location identity number	Yes	No
CountryID	integer	Country identity number. Foriegn key referenced from Dim.Country table.	No	Yes
State	nvarchar(30)	State information	No	No
City	nvarchar(30)	City information	No	No
Latitude	float(19, 12)	Latitude information	No	No
Longitude	float(19, 12)	Longitude information	No	No
GeoNameID	int	Geographical detail about a location	No	No

Column(s) of "DIM.Subject" Table

Name	Datatype	Comment	Is PK	Is FK
SubjectID	integer	Subject Identification	Yes	No
SubjectName	varchar(20)	Subject Name	No	No
Molecule_Team	char(18)	Molecule team information	No	No
CaptureDate	char(18)		No	No

Column(s) of "Fact.InvestigatorPerformance" Table

Name	Datatype	Comment	Is PK	Is FK
InvestigatorID	int	Investigator Identity. The primary key in the table.	Yes	No
SubjectID	integer	Subject Identification. Foreign key from Subject Dimension Table.	Yes	Yes
DataSourceID	int	Data Source Unique Identity Number, apperaing in this table as foreign key	Yes	Yes
LocationID	int	Location identity number	No	Yes
TotalTrials	int	How many trials Investigator & Site participated in this particular indication	No	No
SiteStartupCyleTime	int	Average time for Investigator & Site to open enrollment after the contract signed	No	No
EnrolledINLast16Months	bit	Number of enrollment for the last 16 months	No	No
PatientRandomizedMedian	decimal(19, 12)	Median randamization of patients across all trials for the indication.	No	No
PatientRandomized	int	Average randamization of patients across all trials for the indication.	No	No
PatientRandomizedMaximum	decimal(19, 12)	Maximum randamization of patients across all trials for the indication.	No	No
EnrollmentRateMonthlyMean	decimal(19, 12)		No	No
EnrollmentRateMonthly	int	Enrollment rate per month	No	No
EnrollmentRateMonthlyMeadian	decimal(19, 12)	Median Enrollment Rate per month	No	No
EnrollmentRateMonthlyStandardDev	decimal(19, 12)	Standard Deviation Monthly Data.	No	No

-continued

Column(s) of "Fact.InvestigatorPerformance" Table

Name	Datatype	Comment	Is PK	Is FK
PatientScreened	int	The number of patient screened for this indication.	No	No
ScreenFailureRate	decimal(19, 12)	Percentage of patients that were unable to participate due to failed screening.	No	No
DropoutRate	decimal(19, 12)	Percentage of enrolled patients who dropped out from the trial.	No	No
QueryRate	int	Number of queries per 100 pages of CRFs	No	No
InvestigatorEnrollmentFactor	int	Calculated performence ranking of the Investigators.	No	No
IsActiveFlag	bit		No	No
AuditETLFlag	int		No	No
SiteStartUpCycleTimeStandardDev	decimal(19, 12)	Standard Deviation for Investigator & Site to open enrollment after the contract signed	No	No

Exemplary Country Data Model

[0057] FIG. **12** is an exemplary country data model according to one embodiment of the present invention. In the embodiment shown in FIG. **13**, each country is associated with a unique country identification (CountryID). In this embodiment, the CountryID is associated with other information such as a country name, country abbreviations, population, and/or GPS coordinates. The CountryID for a particular country may be associated with information contained in other tables and/or databases. For example, in the FactPatientPrevalence table shown in FIG. **12**, a CountryID and a subject identification (SubjectID) may be used to determine a prevalence rate, prevalence per population, and incidence per population. As another example, a Country and a SubjectID may be used to query a FactTrialSaturation table to determine an active trial count. The embodiment shown in FIG. **12** depicts numerous other associations between countryIDs and information in other tables and/or databases. Numerous other embodiments are disclosed herein and variations are within the scope of this disclosure.

[0058] Below is a description of the various tables of a country data model according to one embodiment of the present invention:

List of Tables in Country Data Model

Name	Comment
DataSourceOwner	
DIM.Country	Conformed country dimension consisting of ISO-3166 standards and minor manual updates to streamline country presentation data for Semio.
Dim.CountryCycleTime	
Dim.DataSource	
DIM.Subject	Subject table is the dimensional table used to snapshot fact data by molecule team, search phrase, and date.

-continued

List of Tables in Country Data Model

Name	Comment
Fact.CycleTime	Fact table containing cycle time information for a given Country, Subject, and type (CT Materials, throughput, etc)
Fact.MonthlyRecruitment	Fact table containing cycle time information for a given Country, Subject, and type (CT Materials, throughput, etc)
Fact.PatientPrevalence	Contains the prevalence of a particular disease condition (captured in SubjectID) for a given Country (countryID)
Fact.Staffing	Fact table containing staff information for a given Country, Subject, and type.
Fact.TrialSaturation	By Country, By Subject (molecule team + search phrase) - the number of active trials.

Column(s) of "DataSourceOwner" Table

Name	Is PK	Is FK	Comment
DataSourceOwnerID	Yes	No	
DataSourceOwner	No	No	
CreationDate	No	No	
LastModifiedDate	No	No	
LastModifiedByID	No	No	
IsActiveFlag	No	No	

Column(s) of "DIM.Country" Table

Name	Is PK	Is FK	Comment
CountryID	Yes	No	Country Identification Number. The primary key
CountryName	No	No	Country Description
AbbreviationSmall	No	No	Two letter abbreviation name of the country

-continued

Column(s) of "DIM.Country" Table

Name	Is PK	Is FK	Comment
AbbreviationLarge	No	No	Three letters name of the country
IsActiveFlag	No	No	bit indicator for the validity of the record
IsInferredFlag	No	No	
AuditETLID	No	No	Reference to Audit. ExecutionLog key for auditing
FIPS	No	No	Two lettres code for the country
GMI	No	No	Three lettres code for the country
Population	No	No	Population of the country
SQKM	No	No	The total square kilometer of a country
SQMI	No	No	The total miles of a country
Geometery	No	No	Geomatrical information of the country
Landlocked	No	No	The country has a landlocked or not?
CTReferenceCode	No	No	Clinical Trial reference code
Latitude	No	No	Latitude information of the country
Longitude	No	No	Longitude information of the country
ISOName	No	No	

Column(s) of "Dim.CountryCycleTime" Table

Name	Is PK	Is FK	Comment
CountryID	Yes	No	
IterationID	Yes	No	
CycleTimeTypeID	Yes	No	
DataSourceID	Yes	No	
CycleTimeDecimal	No	No	
CycleTimeStandardDev	No	No	
TheoreticalApprovalTimeInDays	No	No	
AverageApprovalTimeInDays	No	No	
SupportingEvidence	No	No	
IsActiveFlag	No	No	
AuditETLID	No	No	
CreatedDate	No	No	
ModifiedDate	No	No	
Modifiedby	No	No	

Column(s) of "Dim.DataSource" Table

Name	Is PK	Is FK	Comment
DataSourceID	Yes	No	Data Source Unique Identity Number
DataSourceName	No	No	Name of the Data Source Provider
DataSourceOwner	No	No	Name of the Data Source Owner
DataSourceDescription	No	No	Description of the Data Source
CreationDate	No	No	The date on which the data is being inserted
LastModifiedDate	No	No	Last modified date
LastModifiedByID	No	No	Last modified by Identity number
IsActiveFlag	No	No	Bit indicator for the validity of the record
DataSourceOwnerID	No	Yes	

Column(s) of "DIM.Subject" Table

Name	Is PK	Is FK	Comment
SubjectID	Yes	No	Subject Identification
SubjectName	No	No	Subject Name
MoleculeTeam	No	No	Molecule team information
CaptureDate	No	No	The date data was inserted in the database
IsLocked	No	No	
IsActiveVariableFlag	No	No	
ProcessFlag	No	No	
SubjectDetailDescription	No	No	

Column(s) of "Fact.CycleTime" Table

Name	Is PK	Is FK	Comment
CountryID	Yes	Yes	Country Identification Number. The primary key
SubjectID	Yes	Yes	Subject Identification
CycleTimeTypeID	Yes	Yes	
DataSourceID	Yes	Yes	Data Source Unique Identity Number
CycleTime	No	No	Cycle time information
SupportiveEvidence	No	No	The xml document containing the supportive information.
IsActiveFlag	No	No	bit indicator for the validity of the record
AuditETLID	No	No	Reference to Audit. ExecutionLog key for auditing
TheoreticalApprovalTimeInDays	No	No	Approval time in days for the country
AverageApprovalTimeInDays	No	No	Average time in days for the country
CyleTimeStandardDev	No	No	Average randamization of patients across all trials for the indication.
IterationID	No	Yes	

Column(s) of "Fact.MonthlyRecruitment" Table

Name	Is PK	Is FK	Comment
CountryID	Yes	Yes	Country Identification Number. The primary key
SubjectID	Yes	Yes	Subject Identification
MonthID	Yes	No	Month Identification
PatientPerSitePerMonth	No	No	Number of patients per site per month
SupportiveEvidence	No	No	The xml document containing the supportive information.
IsActiveFlag	No	No	bit indicator for the validity of the record
AuditETLID	No	No	Reference to Audit. ExecutionLog key for auditing
EnrollmentRateMonthlyStandardDev	No	No	Standard Deviation calculation of Enrollment rate per month
LowerEnrollmentRateMonthlyStandardDev	No	No	Lower side Standard Deviation calculation of Enrollment rate per month

-continued

Column(s) of "Fact.MonthlyRecruitment" Table

Name	Is PK	Is FK	Comment
UpperEnrollmentRateMonthlyStandardDev	No	No	Upper side Standard Deviation calculation of Enrollment rate per month
PatientRandomized	No	No	Randomized Average Number of patients at country level.
DataSourceID	No	Yes	Data Source Unique Identity Number

Column(s) of "Fact.PatientPrevalence" Table

Name	Is PK	Is FK	Comment
CountryID	Yes	Yes	Country Identification Number. The primary key
SubjectID	Yes	Yes	Subject Identification
PrevalencePer	No	No	The column contains the number of patient for calculation
PrervalenceRate	No	No	The prevalence rate calculated number
SupportingEvidence	No	No	The xml document containing the supportive information.
IsActiveFlag	No	No	bit indicator for the validity of the record
AuditETLID	No	No	Reference to Audit. ExecutionLog key for auditing
DataSourceID	No	Yes	Data Source Unique Identity Number
PrevalencePerPopulation	No	No	
IncidencePer	No	No	
IncidencePerPopulation	No	No	
IncidenceSupportingEvidence	No	No	

Column(s) of "Fact.Staffing" Table

Name	Is PK	Is FK	Comment
CountryID	Yes	Yes	Country Identification Number. The primary key
SubjectID	Yes	Yes	Subject Identification
CRANumber	No	No	Total number of CRA in a country
SupportiveEvidence	No	No	XML document containing evidentiary details of the findings in the fact.
IsActiveFlag	No	No	bit indicator for the validity of the record
AuditETLID	No	No	Reference to Audit. ExecutionLog key for auditing
StaffTotal	No	No	Total number of staff
DataSourceID	No	Yes	Data Source Unique Identity Number

Column(s) of "Fact.TrialSaturation" Table

Name	Is PK	Is FK	Comment
CountryID	Yes	Yes	Country Identification Number. The primary key
SubjectID	Yes	Yes	Subject Identification
ActiveTrialCount	No	No	Number of Active Trail information
SupportiveEvidence	No	No	The xml document containing the supportive information.
IsActiveFlag	No	No	bit indicator for the validity of the record
AuditETLID	No	No	Reference to Audit. ExecutionLog key for auditing
DataSourceID	No	Yes	Data Source Unique Identity Number

Exemplary Patient Data Model

[0059] FIG. **13** is an exemplary patient data model according to one embodiment of the present invention. In the embodiment shown in FIG. **13**, a FactPatient table stores information associated with various patients. For example, patients in the FactPatient table may be assigned a unique patient identification number (PatientID). In this embodiment, the Patient ID can be associated with other information such as the patient's age, gender, year of birth, and/or other information. The PatientID may also be associated with information contained in the other tables and/or databases. For example, a Patient ID may be associated with a location identification (LocationID). In this embodiment, the LocationID for the patient corresponds with a location identification of a separate table (DimLocation). The LocationID) for the patient can be used to determine information such as a city, a state, and/or GPS coordinates associated with the PatientID based on the LocationID. As another example, a SubjectID associated with a particular PatientID may be used to determine a subject's name, molecule team, capture date, and/or other information contained in another table and/or database having a corresponding SubjectID. The embodiment shown in FIG. **13**, depicts numerous other associations between information corresponding to PatientIDs and information in other tables and/or databases. Furthermore, numerous other embodiments are disclosed herein and variations are within the scope of this disclosure. Numerous other data models and variations of the data models described herein are likewise within the scope of this disclosure.

[0060] Below is a description of the various tables of a patient data model according to one embodiment of the present invention:

List of Tables for Patient Data Model

Name	Comment
DataSourceOwner	
Dim.ConcomitantMedication	ConcomitantMedicationn dimension table contains name of the medication, class information.
Dim.DataSource	
Dim.Ethnicity	Ethnicity dimension table containsethnicity information about patient

-continued

List of Tables for Patient Data Model

Name	Comment
Dim.Location	Location dimension table contains location information about country, state, city, latitude, and longitude.
DIM.Subject	Subject table is the dimensional table used to snapshot fact data by molecule team, search phrase, and date.
Fact.Patient	Fact table containing information for a patient, Subject
Fact.PatientCo-MorbidityCondition	
Fact.PatientCon-comitantMedication	
ICD9.Codes	
Medication	
TreatmentType	

Column(s) of "DataSourceOwner" Table

Name	Datatype	Comment	Is PK	Is FK
DataSourceOwnerID	int		Yes	No
DataSourceOwner	nvarchar(50)		No	No
CreationDate	datetime		No	No
LastModifiedDate	datetime		No	No
LastModifiedByID	varbinary(85)		No	No
IsActiveFlag	bit		No	No

Column(s) of "Dim.ConcomitantMedication" Table

Name	Datatype	Comment	Is PK	Is FK
Concomi-tantMedicationID	int	Concomitant Medication Identity. The Primary Key of the table	Yes	No
MedicationName	nvarchar(50)	Name of the medication	No	No
MedicationClass	nvarchar(50)	Medical Class information.	No	No
DDI	int	Drug Index	No	No
NDC	bigint	Drug Index	No	No
GPI	bigint	Drug Index	No	No
DataSourceID	int	Data Source Identity	No	No

Column(s) of "Dim.DataSource" Table

Name	Datatype	Comment	Is PK	Is FK
DataSourceID	int	Data Source Unique Identity Number	Yes	No
DataSourceName	nvarchar(50)	Name of the Data Source Provider	No	No
DataSourceDe-scription	nvarchar(255)	Description of the Data Source	No	No
FileDate	datetime		No	No
CreationDate	datetime	The date on which the data is being inserted	No	No

-continued

Column(s) of "Dim.DataSource" Table

Name	Datatype	Comment	Is PK	Is FK
LastModifiedDate	datetime	Last modified date	No	No
LastModifiedByID	varbinary(85)	Last modified by Identity number	No	No
IsActiveFlag	bit	Bit indicator for the validity of the record	No	No
DataSourceOwnerID	int	Data Source Owner description	No	Yes

Column(s) of "Dim.Ethnicity" Table

Name	Datatype	Comment	Is PK	Is FK
EthnicityID	int	Ethnicity Identity. Primary key of the table	Yes	No
Ethnicity	nvarchar(50)	Details about ethnicity	No	No
NISNumber	int		No	No
SelfReferenceID	int		No	No

Column(s) of "Dim.Location" Table

Name	Datatype	Comment	Is PK	Is FK
LocationID	int	Location identity number	Yes	No
CountryID	int	Country identity number. Foriegn key referenced from Dim.Country table.	No	No
State	nvarchar(30)	State information	No	No
City	nvarchar(30)	City information	No	No
AsciiName	Nvarchar(200)		No	No
Latitude	float(19, 12)	Latitude information	No	No
Longitude	float(19, 12)	Longitude information	No	No
GeoNameID	int	Geographical detail about a location	No	No
SelfReference	int	Self referenced number	No	No

Column(s) of "DIM.Subject" Table

Name	Datatype	Comment	Is PK	Is FK
SubjectID	integer	Subject Identification	Yes	No
SubjectName	varchar(20)	Subject Name	No	No
MoleculeTeam	char(18)	Molecule team information	No	No
CaptureDate	char(18)	The date on which the record was captured.	No	No
IsLocked	bit		No	No
IsActiveVariableFlag	bit		No	No
ProcessFlag	bit		No	No
SubjectDetailDescription	nvarchar(max)		No	No

Column(s) of "Fact.Patient" Table

Name	Datatype	Comment	Is PK	Is FK
PatientID	int	Patient Identification Number	Yes	No
SubjectID	integer	Subject Identification	No	Yes
EthnicityID	int	Ethnicity Identity. Primary key of the table	No	Yes
LocationID	int	Location identity number	No	Yes
PatientSourceID	int	This field indicates to the source for the patient information.	No	No
DataSourceID	int	Data Source Identity number indicating the source of the data	No	Yes
Gender	varchar( )	Patient Gender Information	No	No
Age	numeric(,)	Patient Age	No	No
BirthYear	int	Date of Birth Year	No	No
SourcePatientID	int	Source Patient Identity	No	No
Ethnicity	nvarchar(50)	Information about Ethnicity	No	No
AuditETLID	int	Reference to Audit. ExecutionLog key for auditing	No	No
IsActiveFlag	bit	Bit indicator for the validity of the record	No	No
BMI	decimal(19, 12)	Body Mass Index	No	No
CreatinineValue	decimal(19, 12)		No	No
eGERValue	decimal(19, 12)		No	No
ProteinCreatinineRatio	decimal(19, 12)		No	No

Column(s) of "Fact.PatientCoMorbidityCondition" Table

Name	Datatype	Comment	Is PK	Is FK
PatientCoMorbidityCondtionID	int	Patient CoMorbidity Condition Unique Identity Number. The Primary Key of the	Yes	No
PatientID	int	Patient Identification Number. Foreign Key from Fact.PatientType	No	Yes
CodeID	int		No	No
SubjectID	integer	Subject Identification Number. Foreign Key from Dim.Subject.	No	Yes
MedcinID	int	Medication Identity Number	No	No
Type	nvarchar(55)		No	No
Category	nvarchar(55)		No	No
Status	nvarchar(55)		No	No
OnSetDate	datetime		No	No
CreationDate	datetime		No	No
LastModifiedDate	datetime		No	No
LastModifiedByID	varbinary(85)		No	No
IsActiveFlag	bit		No	No
DataSourceID	int	Data Source Unique Identity Number	No	Yes

-continued

Column(s) of "Fact.PatientCoMorbidityCondition" Table

Name	Datatype	Comment	Is PK	Is FK
ICD9CodeID	int	Code Identity	No	Yes
DiagnosisTypeID	int		No	No

Column(s) of "Fact.PatientConcomitantMedication" Table

Name	Datatype	Comment	Is PK	Is FK
PatientConcomitantMedicationID	int	PatientConcomitantMedicationID is the unique primary key to the table	Yes	No
PatientID	int	Patient Identification Number	No	Yes
ConcomitantMedicationID	int	Concomitant Medication Identity. The Primary Key of the table	No	Yes
SubjectID	integer	Subject Identification	No	Yes
Dose	nvarchar(55)	Dose of the Medication	No	No
Strength	nvarchar(55)	Strength of the Medication	No	No
Form	nvarchar(55)	Form of the Medication	No	No
Units	nvarchar(55)	Unit of the Medication	No	No
Quantity	nvarchar(55)	Quantity of the Medication	No	No
Status	nvarchar(55)	Status of the Medication	No	No
CreationDate	datetime	Bit indicator for the validity of the record	No	No
LastModifiedDate	datetime		No	No
LastModifiedByID	varbinary(85)		No	No
IsActiveFlag	bit		No	No
DataSourceID	int	Data Source Unique Identity Number	No	Yes
MedicationID	int		No	Yes
TreatmentTypeID	int		No	Yes

Column(s) of "ICD9.Codes" Table

Name	Datatype	Comment	Is PK	Is FK
ICD9CodeID	int		Yes	No
IndicationID	int		No	No
IndicationGroupID	int		No	No
Code	varchar(20)		No	No
Description	varchar(255)		No	No
IndicationGroupCode	varchar(20)		No	No
ParentCodeID	int		No	No
DataSourceID	int	Data Source Identity number indicating the source of the data	No	No

Column(s) of "Medication" Table

Name	Datatype	Comment	Is PK	Is FK
MedicationiD	int		Yes	No
MedicationName	nvarchar(75)		No	No
DataSourceiD	int		No	No

Column(s) of "TreatmentType" Table

Name	Datatype	Comment	Is PK	Is FK
TreatmentTypeiD	int		Yes	No
TreatmeniTvoeDescriotion	nvarchar(50)		No	No
CreationDate	date		No	No

Advantages

[0061] Embodiments of the present invention provide many advantages over conventional methods of predicting the enrollment for clinical trials. For example, embodiments of the present invention allow subject identifications (Os) to be created through one or more user interfaces. In one embodiment, a user can create one or more subject IDs without technical expertise. For example, using one or more user interfaces, a user can create a new subject by entering or selecting a subject name, molecule team, and/or a capture date. A unique subject identification may be dynamically created for the subject In another embodiment, a user can update an existing subject. For example, a user may be able to add or update a capture date or other information associated with a particular subject ID. Based at least in part on the capture data, data from various tables and/or databases associated with a subject ID may be limited. For example, the capture date may provide a static point in time for which information contained in the tables and/or databases is available. Thus, if the information for a particular table and/or database specifies that the information is before the capture date, then the information is available to the subject identification. Alternatively, if the information for a particular table and/or database specifies that the information is after the capture date, then this information may not be available to the subject identification.

[0062] Embodiments of the present invention provide one or more core domains of information that may be used for analysis of a clinical trial plan. For example, patient domains, country domains, and/or investigator domains of information can be used according to one embodiment. In some embodiments, one or more core domains are built generically such that the system can manage data sources that are unknown at the time the core domain is created. In this way, using a generic structure that supports a domain, additional data sources can be added and/or updated as additional inthrmation and/or data sources become available.

[0063] Subject identifications may be associated with at least a portion of the information for one or more domains. For example, subject identifications may be associated with information contained in a patient domain, a country domain, an investigator domain, and/or other domains or data sources.

[0064] Once various parameters are chosen and associations between subject identifications and information in the domains have been created, embodiments of the present invention are able to take a mathematical approach to analyzing and presenting data regarding the actual investigators and investigation sites. The embodiments can then create graphical representations, e.g., line graphs that display information, such as predictions for likely scenarios based on average performance as well as best and worst-case scenarios based on outlier data.

General

[0065] Numerous specific details are set forth herein to provide a thorough understanding of the claimed subject matter. However, those skilled in the art will understand that the claimed subject matter may be practiced without these specific details. In other instances, methods, apparatuses or systems that would be known by one of ordinary skill have not been described in detail so as not to Obscure claimed subject matter.

[0066] Some portions are presented in terms of algorithms or symbolic representations of operations on data bits or binary digital signals stored within a computing system memory, such as a computer memory. These algorithmic descriptions or representations are examples of techniques used by those of ordinary skill in the data processing arts to convey the substance of their work to others skilled in the art. An algorithm is a self-consistent sequence of operations or similar processing leading to a desired result. In this context, operations or processing involves physical manipulation of physical quantities. Typically, although not necessarily, such quantities may take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared or otherwise manipulated. It has proven convenient at times, principally for reasons of common usage, to refer to such signals as bits, data, values, elements, symbols, characters, terms, numbers, numerals or the like. It should be understood, however, that all of these and similar terms are to be associated with appropriate physical quantities and are merely convenient labels. Unless specifically stated otherwise, it is appreciated that throughout this specification discussions utilizing terms such as "processing," "computing," "calculating," "determining," and "identifying" or the like refer to actions or processes of a computing device, such as one or more computers or a similar electronic computing device or devices, that manipulate or transform data represented as physical electronic or magnetic quantities within memories, registers, or other information storage devices, transmission devices, or display devices of the computing platform.

[0067] The system or systems discussed herein are not limited to any particular hardware architecture or configuration. A computing device can include any suitable arrangement of components that provide a result conditioned on one or more inputs. Suitable computing devices include multipurpose microprocessor-based computer systems accessing stored software that programs or configures the computing system from a general purpose computing apparatus to a specialized computing apparatus implementing one or more embodiments of the present subject matter. Any suitable programming, scripting, or other type of language or combinations of languages may be used to implement the teachings contained herein in software to be used in programming or configuring a computing device.

[0068] Embodiments of the methods disclosed herein may be performed in the operation of such computing devices. The order of the blocks presented in the examples above can be varied—for example, blocks can be re-ordered, combined, and/or broken into sub-blocks. Certain blocks or processes can be performed in parallel.

[0069] The use of "adapted to" or "configured to" herein is meant as open and inclusive language that does not foreclose devices adapted to or configured to perform additional tasks or steps. Additionally, the use of "based on" is meant to be open and inclusive, in that a process, step, calculation, or other action "based on" one or more recited conditions or values may, in practice, be based on additional conditions or values beyond those recited. Headings, lists, and numbering included herein are for ease of explanation only and are not meant to be limiting.

[0070] While the present subject matter has been described in detail with respect to specific embodiments thereof, it will be appreciated that those skilled in the art, upon attaining an understanding of the foregoing may readily produce alterations to, variations of, and equivalents to such embodiments. Accordingly, it should be understood that the present disclosure has been presented for purposes of example rather than limitation, and does not preclude inclusion of such modifications, variations and/or additions to the present subject matter as would be readily apparent to one of ordinary skill in the art.

That which is claimed is:

1. A method comprising:

receiving selection of a subject identification, the subject identification corresponding with at least a subject name;

receiving selection of a patient parameter, the patient parameter corresponding with at least one of a physical characteristic or an illness;

receiving selection of an investigator parameter, the investigator parameter corresponding with at least one of a trial, a phase, or a medical indicator;

receiving selection of a geographic parameter, the geographic parameter corresponding with at least one of a geographic location or a geographic statistic;

dynamically creating associations between the subject identification and the patient parameter, the investigator parameter, and the geographic parameter; and

determining at least one potential site for a clinical trial based at least in part on the dynamically created associations.

2. The method of claim **1**, further comprising:

conducting the clinical trial based at least in part on the determined at least one potential site.

3. The method of claim **1**, where receiving selection of the subject identification comprises:

receiving selection of the subject name from a plurality of predefined subject names.

4. The method of claim **1**, where receiving selection of the subject identification comprises:

receiving a first input, the first input indicating the subject name;

receiving a second input, the second input indicating a molecule team; and

dynamically creating the subject identification.

5. The method of claim **1**, wherein receiving selection of the patient parameter comprises receiving selection of at least one medical code, wherein the at least one medical code comprises at least one ICD9 code or ICD10 code.

6. The method of claim **1**, further comprising:

in response to receiving the selection of the patient parameter, displaying a number of patients having the selected patient parameter.

7. The method of claim **6**, wherein the number of patients having the selected patient parameter is selected from a plurality of patient databases, each patient database in the plurality of patient databases corresponding to a same generic patient data model.

8. The method of claim **1**, wherein dynamically creating associations between the subject identification and the patient parameter, the investigator parameter, and the country parameter comprises:

creating at least one association between the subject identification and at least one patient database comprising the patient parameter;

creating at least one association between the subject identification and at least one investigator database comprising the investigator parameter; and

creating at least one association between the subject identification and at least one country database comprising the country parameter.

9. The method of claim **8**,

wherein each patient database corresponds to a same patient data model,

wherein each investigator database corresponds to a same investigator data model, and

wherein each country database corresponds to a same country data model.

10. The method of claim **8**, further comprising:

receiving selection of a capture date; and

updating the subject identification such that only information contained in the at least one patient database, the at least one investigator database, and the at least one country database on or before the capture date is available for the selected subject identification.

11. The method of claim **1**, further comprising:

creating a graphical representation based at least in part on the dynamically created associations, the graphical representation providing a prediction for one or more scenarios.

12. The method of claim **11**, wherein the prediction corresponds to at least one of a likely scenario based at least on part on average performance, a best-case scenario, or a worst-case scenario.

13. A computer-readable medium comprising program code for:

receiving selection of a subject identification, the subject identification corresponding with at least a subject name;

receiving selection of a patient parameter, the patient parameter corresponding with at least one of a physical characteristic or an illness;

receiving selection of an investigator parameter, the investigator parameter corresponding with at least one of a trial, a phase, or a medical indicator;

receiving selection of a geographic parameter, the geographic parameter corresponding with at least one of a geographic location or a geographic statistic;

dynamically creating associations between the subject identification and the patient parameter, the investigator parameter, and the geographic parameter; and

sending information corresponding to at least one of the dynamically created associations to a clinical trial analysis application for conducting a clinical trial.

14. The computer-readable medium of claim **13**, further comprising program code for:

receiving a capture date, wherein only information contained in at least one patient database, at least one investigator database, and at least one country database as of the capture date is available for the subject identification.

15. The computer-readable medium of claim **13**, further comprising program code for:

creating a graphical representation for the clinical trial based at least in part on the dynamically created associations.

16. The computer-readable medium of claim **13**, further comprising program code for:

associating a capture date with the subject identification such that data on or before the capture date can be maintained in a database while allowing the database to continue to receive additional data after the capture date.

17. A system, comprising:

a plurality of databases;

an electronic device comprising:

an input device;

a display; and

a processor in communication with the input device, the display, the plurality of databases, the processor configured for:

receiving selection of a subject identification, the subject identification corresponding with at least a subject name;

receiving selection of a patient parameter, the patient parameter corresponding with at least one of a physical characteristic or an illness;

receiving selection of an investigator parameter, the investigator parameter corresponding with at least one of a trial, a phase, or a medical indicator;

receiving selection of a geographic parameter, the geographic parameter corresponding with at least one of a geographic location or a geographic statistic;

dynamically creating associations between the subject identification and the plurality of databases for use in a clinical trial analysis application, the associations comprising a first association between the subject identification and the patient parameter, a second association between the subject identification and the investigator parameter, and a third association between the subject identification and the country parameter.

18. The system of claim **17**, further comprising:

a network; and

a server comprising:

a memory, wherein the memory comprises program code for the clinical trial analysis application;

a second processor, the second processor in communication with the memory, the second processor configured for:

executing the program code for the clinical trial analysis application;

receiving information corresponding to at least one of the dynamically created associations from the electronic device through the network; and

conducting a clinical trial using at least the clinical trial analysis application and the received information.

19. The system of claim **17**,

wherein the plurality of databases comprises a patient database, an investigator database, and a country database, and

wherein the processor is further configured for:

querying the patient database with the patient parameter;

querying the investigator database with the investigator parameter; and

querying the country database with the country parameter;

20. The system of claim **17**, wherein the processor is further configured for:

receiving a capture date; and

filtering information in the plurality of databases such that only information in the plurality of databases on or before the capture date is available to the subject identification and such that new information can be added to the plurality of databases after the capture date and the new information is not available to the subject identification.

* * * * *