SUGGESTING POTENTIAL CUSTODIANS FOR CASES IN AN ENTERPRISE-WIDE ELECTRONIC DISCOVERY SYSTEM

CLAIM OF PRIORITY UNDER 35 U.S.C. §119

[0001] The present Application for Patent claims priority to Provisional Application No. 61/164,276 entitled "Electronic Discovery System" filed March 27, 2009, and assigned to the assignee hereof and hereby expressly incorporated by reference herein.

FIELD

[0002] In general, embodiments of the invention relate to methods, systems, apparatus and computer program products for electronic discovery and, more particularly, determining suggestions for potential custodians to add to a case in an enterprise-wide electronic discovery system.

BACKGROUND

[0003] Electronic discovery, commonly referred to as e-discovery, refers to any process in which electronic data is sought, located, secured and searched with the intent of using it as evidence in a legal proceeding, an audit, a securities investigation, a forensics investigation or the like. E-discovery can be carried out offline on a particular computer or it can be accomplished in a network environment.

[0004] The nature of digital data makes it extremely well-suited for investigation. In particular, digital data can be electronically searched with ease, whereas paper documents must be scrutinized manually. Furthermore, digital data is difficult or impossible to completely destroy, particularly if the data is stored in a network environment. This is because the data appears on multiple hard drives, and because digital files, even if deleted, generally can be undeleted. In fact, the only reliable means of destroying digital data is to physically destroy any and all hard drives where it is stored.

[0005] In the process of electronic discovery, data of all types can serve as evidence. This can include text, image, calendar event data, databases, spreadsheets, audio files, multimedia files, web sites and computer programs. Electronic mail (*i.e.*, e-mail) can be an especially valuable source of evidence in civil or criminal litigation, because people are often

less careful in these exchanges than in hard copy correspondence such as written memos or postal letters.

[0006] E-discovery is an evolving field that goes far beyond mere technology. It gives rise to multiple issues, many of which have yet to be resolved. For example, identifying data required to satisfy a given discovery request, locating the appropriate set of data that has been identified, and retrieving the data once it has been identified and located all pose problems in and of themselves. This is especially evident if the data that is being identified, located and retrieved comes from an evolving or disparate enterprise, such as a corporation that has experienced mergers, acquisitions, downsizing and the like. Mergers and acquisitions mean that the technology infrastructure across the enterprise may vary, at least in the interim. However, e-discovery must be able locate and retrieve data from these disparate technology infrastructure in a timely fashion, sometimes within days of when the merger/acquisition occurs.

[0007] In addition to identifying, locating and retrieving digital data, the most critical part of any electronic discovery is the preservation of data, which involves maintaining an original source copy and storing it for preservation purposes or furthering processing. This too becomes a daunting task for the enterprise system that encompasses a myriad of different technology infrastructures and the like. Therefore, a need exists to improve the identification, location, retrieval and preservation processes, especially in instances in which the enterprise system includes disparate technology infrastructures and the like.

[0008] As previously noted, e-discovery, as opposed as conventional discovery of printed materials, provides for the ability to filter or search the data so as to reduce the volume of data to only that which is relevant to the request. Such searching is typically accomplished by determining a specific date range for the request, providing key words relevant to the case and the like. Improvements in the area of searching are greatly in need to further add efficiency to the overall e-discovery process.

[0009] Once data has been retrieved, preserved and, in some instances, searched the electronic data may be reviewed by the requesting entry, such as a law firm, securities commission or the like. While large requests are generally suited for online review, the manner in which the data is presented for review adds efficiency to the review process and ultimately drives the cost of the review process. Therefore, improvements in the manner in which data is presented for review are also desirable as a means of increasing efficiency and reducing costs.

[0010] Lastly, once the digital data has been reviewed, data identified as relevant may need to be produced in a tangible format for further analysis or legal evidentiary purposes. The produced documents must be properly identified and include necessary redactions and confidentiality markings.

[0011] Up until now, e-discovery management has been conducted on a case-by-case basis, meaning all tasking and workflow related to the e-discovery is based at the case level. Such management does not allow for finer granularity in the management of a case or for links to exist between different cases for the purpose of leveraging the e-discovery related to one case to another new or pre-existing case. Therefore, a need exists to improve the manner in which cases are managed and, in particular, how tasking and workflow are managed depending on case requirements and the like.

[0012] At the onset of a case, electronic discovery associates have the daunting task of trying to determine who is or might be associated with a case. This is a difficult and time consuming endeavor that requires the electronic discovery associates to perform research to determine who is or might be associated with a case. The research requires the electronic discovery associate to interview individuals already identified as being associated with a case and/or access enterprise-related databases, such as personnel databases or the like, to determine if someone is or might be associated with a case. This research process is not foolproof and, all too often, during the actual discovery phase of a case it is determined that one or more individuals associated with a case have not been added to the case. This is because of oversights or limitations in determining who is or might be associated with a case during the research process.

[0013] Therefore, a need exists to automate the process whereby individuals within an enterprise system can be readily identified as being potentially associated with a case. The desired process should take into account who is already added to a case and determine who else is or might be associated with the case based on their association, during the timeframe of the case, with those previously added individuals. The desired process and system eliminates the need to manually research who is or might be associated with a case and, in doing so, lessens the likelihood of a failure to associate/add individuals to a case due to human-error or limitations in manual research.

SUMMARY

[0014] The following presents a simplified summary of one or more embodiments in order to provide a basic understanding of such embodiments. This summary is not an extensive overview of all contemplated embodiments, and is intended to neither identify key or critical elements of all embodiments, nor delineate the scope of any or all embodiments. Its sole purpose is to present some concepts of one or more embodiments in a simplified form as a prelude to the more detailed description that is presented later.

[0015] Embodiments of the present invention relate to systems, apparatus, methods, and computer program products for electronic discovery and, in particular, improvements in electronic discovery that allow for electronic discovery to be efficiently and cost-effectively employed across a diverse enterprise. In particular, embodiments herein described related to systems, apparatus, methods and computer program products for automatically determining suggestions for who is or might be associated with a case (*i.e.*, potential custodians or individuals of interest in the specified case) based on who is already been added to the case and taking into account the timeframe of the case.

[0016] Thus, present embodiments provide for determining suggestions for potential custodians to be added to an electronic discovery case. The determination is based on a comparison of data, such as personnel data or like, associated with individuals already added to the case and all other individuals in the enterprise, who form the group of potential custodians. In addition, the comparison of data occurs over the specified timeframe defined by the case to insure that the suggested potential custodians are within the confines to the timeframe of the case. Once the suggested potential custodians have been determined they are presented to the electronic discovery associate charged with adding potential custodians to the case and the associate can make a final determination as to whether to add or not add the suggested potential custodian to the case.

[0017] Additional embodiments of the invention provide for determining preservation notice recipients based on a comparison of data, such as personnel data or the like, associated with individuals that have already been identified to receive the preservation notice and other custodians and/or potential custodians. In addition, the determination takes into account the preservation notice history of the custodian and/or potential custodian being considered as a recipient of preservation notice of a specific preservation notice.

[0018] The present invention eliminates the need for the electronic discovery associate to perform manual research to determine which individuals within the enterprise to add to an electronic discovery case or, subsequently, which custodians to send preservation notices. In addition, the present invention makes it less likely for omissions to occur at the initial identification of potential custodian phase or intended preservation notice recipients. Thus, by automating the process a time savings is realized and the possibility of omissions, oversights and the like in identifying potential custodians and/or preservation notice recipients are lessened; adding to the overall reliability of the identification of potential custodians and preservation notice recipients.

[0019] A method for suggesting potential custodians for a case in an enterprise-wide electronic discovery system defines specific embodiments of the invention. The method includes receiving, at computing device, first inputs that define a case including a timeframe associated with the case and receiving, at a computing device, second inputs that add one or more potential custodians to the case. The method further includes determining, via computing device processor, a plurality of suggested potential custodians. The determination is based on association with the one or more previously added potential custodians and the timeframe associated with the case. Additionally, the method includes providing, at a computing device, an output of the determined plurality of suggested potential custodians. In further specific embodiments, the method includes receiving, at a computing device, third inputs that add one or more of the suggested potential custodians to the case.

[0020] According to specific embodiments of the method, determining may further entail determining, via the computing device processor, at least one match between predetermined data associated with at least one added potential custodian and a suggested potential custodian. The predetermined data may include, but is not limited to, employment chain-of-command, employment locale, employment classification, data source accessibility and/or generated data content. This determining may further include accessing, via a computing device processor, one or more enterprise databases to determine the at least one match between the predetermined data associated with at least one added potential custodian and a suggested potential custodian.

[0021] In other specific embodiments of the method, in which matching of predefined data is included, a predefined threshold of matches may be defined, such that determination of a suggested potential custodian may entail determining if the threshold of matches has been met.

In such embodiments, providing may further entail providing, at the computing device, an output of the determined plurality of suggested potential custodians that meet the predefined threshold of matches.

[0022] In still further specific embodiments of the method, in which matching of predefined data is included, determining may further include applying a predetermined relevancy weight value to each match of the predetermined data and summing or otherwise calculating the predetermined relevancy weight values to result in a suggested potential custodian score. In such embodiments, providing an output of the determined plurality of suggested potential custodians may further include providing a ranked listing of the suggested potential custodians, wherein the ranked listing is ranked based on the suggested potential custodian score.

[0023] In other specific embodiments of the method, in which matching of predefined data is required, clustering techniques may be implemented to determine suggested potential custodians and to refine the output of suggested potential custodians based on more potential custodians being added to the case.

[0024] In other specific embodiments, the method may include determining, at a computing device, one or more suggested recipients of a preservation notice from amongst potential custodians previously added to the case. The determination is based on an association with the one or more previously identified recipients of the preservation notice and previous preservation notice history of the one or more suggested recipients. In such embodiments, determining one or more suggested recipients may further include determining, via the computing device processor, at least one match between predetermined data associated with at least one of the previous identified recipients and a suggested recipient. The predetermined data associated with the at least one of the previous identified recipients and the suggested recipient may include, but is not limited to, employment chain-of-command, employment locale, employment classification, data source accessibility and/or generated data content. In related embodiments, determining one or more suggested recipients further includes determining, via the computing device processor, that a predefined threshold of matches exist between the predetermined data associated with the at least one of the previous identified recipients and the suggested recipient. In still other related embodiments, determining one or more suggested recipients, further includes applying a predetermined relevancy weight value to each match of

the predetermined data and summing or otherwise calculating the predetermined relevancy weight values to result in a suggested potential preservation notice recipient score.

[0025] An apparatus for suggesting potential custodians for a case in an enterprise-wide electronic discovery system provides for further embodiments of the invention. The apparatus, which may include multiple devices, includes a computing platform including a memory and at least one processor. The apparatus further includes an electronic discovery manager application stored in the memory, executable by the at one processor and configured to receive first inputs that define a case and include a timeframe associated with the case and receive second inputs that add one or more potential custodians to the case. The apparatus additionally includes a suggested potential custodian application stored in the memory, executable by the at least one processor and configured to determine a plurality of suggested potential custodians. The determination is based on an association with the one or more previously added potential custodians and the timeframe associated with the case. The suggested potential custodian application is further configured to provide an output of the determined plurality of suggested potential custodians.

[0026] In other specific embodiments, the apparatus may include a case database located in the memory and configured to store case-related data including the timeframe associated with the case and a custodian database located in the memory and configured to store potential custodian and actual custodian-related data. The case database is linked to the custodian database such that adding one or more potential custodians to the case is reflected in respective case-related data in the custodian database and the potential custodian-related data in the custodian database.

[0027] In other specific embodiments of the apparatus, the suggested potential custodian application is further configured to determine at least one match between predetermined data associated with at least one added potential custodian and a suggested potential custodian. The predetermined data includes, but is not limited to, employment chain-of-command, employment locale, employment classification, data source accessibility and/or data content generated by potential custodians. Thus, in such embodiments, the suggested potential custodian application is further configured to access one or more enterprise databases to determine the at least one match between the predetermined data associated with at least one added potential custodian and a suggested potential custodian.

[0028] In further specific embodiments of the apparatus, in which the suggested potential custodian application is configured to determine matches in predetermined data between previously added potential custodians and suggested potential custodians, the suggested potential custodian application may be further configured to determine that a predefined threshold of matches exist between the predetermined data associated with at least one added potential custodian and a suggested potential custodian. Further, in such embodiments, the suggested potential custodian application may be further configured to provide the output of the determined plurality of suggested potential custodians that meet the predefined threshold of matches.

[0029] Moreover, in other specific embodiments of the apparatus, in which the suggested potential custodian application is configured to determine matches in predetermined data between previously added potential custodians and suggested potential custodians, the suggested potential custodian application is further configured to apply a predetermined relevancy weight value to each match of the predetermined data and sum the predetermined relevancy weight values to result in a suggested potential custodian score. In such embodiments, the suggested potential custodian application may be further configured to provide the output of the determined plurality of suggested potential custodians, wherein the output comprises a ranked listing of the suggested potential custodians, and wherein the ranked listing is ranked based on the suggested potential custodian score.

[0030] In further embodiments, the apparatus may include a suggested preservation notice recipient application stored in the memory, executable by the processor and configured to determine one or more suggested recipients of a preservation notice from amongst potential custodians previously added to the case. The determination is based on an association with the one or more previously identified recipients of the preservation notice and previous preservation notice history of the one or more suggested recipients.

[0031] A computer program product that includes a computer-readable medium defines yet another embodiment of the invention. The computer-readable medium includes a first set of codes for causing a computer to receive first inputs that define a case including a timeframe associated with the case and a second set of codes for causing a computer to receive second inputs that add one or more potential custodians to the case. The computer-readable medium additionally includes a third set of codes for causing a computer to determine a plurality of

suggested potential custodians. The determination is based on association with the one or more previously added potential custodians and the timeframe associated with the case. Additionally, the computer-readable medium includes a fourth set of codes for causing a computer to provide an output of the determined plurality of suggested potential custodians.

[0032] Thus, further details are provided below for systems, apparatus, methods and computer program products for determining suggestions for potential custodians to be added to an electronic discovery case. As noted, the determination is based on a comparison of data, such as personnel data or like, associated with individuals already added to the case and all other individuals in the enterprise, who form the group of potential custodians. In addition, the comparison of data occurs over the specified timeframe defined by the case to insure that the suggested potential custodians are within the confines to the timeframe of the case.

[0033] To the accomplishment of the foregoing and related ends, the one or more embodiments comprise the features hereinafter fully described and particularly pointed out in the claims. The following description and the annexed drawings set forth in detail certain illustrative features of the one or more embodiments. These features are indicative, however, of but a few of the various ways in which the principles of various embodiments may be employed, and this description is intended to include all such embodiments and their equivalents.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] Having thus described embodiments of the invention in general terms, reference will now be made to the accompanying drawings, which are not necessarily drawn to scale, and wherein:

[0035] **FIG. 1** is schematic diagram of an apparatus configured for determining suggestive potential custodians in an electronic discovery system, in accordance with embodiments of the present invention;

[0036] **FIG. 2** is a detailed block diagram of an apparatus configured for determining suggestive potential custodians in an electronic discovery system, in accordance with embodiments of the present invention;

[0037] **FIG. 3** is a flow diagram of a method for determining suggestive potential custodians in an electronic discovery system, in accordance with embodiments of the present invention;

[0038] **FIG. 4** is a block diagram of an enterprise-wide electronic discovery system highlighting computing device monitoring, in accordance with embodiments of the present invention;

[0039] **FIG. 5** is a block diagram of an electronic discovery manager server, in accordance with embodiments of the present invention;

[0040] **FIG. 6** is a block diagram of a database server in an electronic discovery system, in accordance with an embodiment of the present invention;

[0041] **FIG. 7** is a block diagram of a collection server in an electronic discovery system, in accordance with an embodiment of the present invention;

[0042] **FIG. 8** is block diagram illustrating electronic discovery management structure, in accordance with an embodiment of the invention;

[0043] **FIG. 9** is a flow diagram of a method for initiating a case or matter including creating search terms, creating and sending preservation notices, sending reminder notices and creating and sending surveys to custodians, in accordance with embodiments of the present invention;

[0044] **FIG. 10** is a flow diagram of a method for custodian management in an electronic discovery system, in accordance with an embodiment of the present invention; and

[0045] **FIGS. 11** and **12** are flow diagrams of methods for harvesting different data types in an electronic discovery system, in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

[0046] Embodiments of the present invention now may be described more fully hereinafter with reference to the accompanying drawings, in which some, but not all, embodiments of the invention are shown. Indeed, the invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure may satisfy applicable legal requirements. Like numbers refer to like elements throughout.

[0047] As may be appreciated by one of skill in the art, the present invention may be embodied as a method, system, computer program product, or a combination of the foregoing. Accordingly, the present invention may take the form of an entirely software embodiment (including firmware, resident software, micro-code, *etc.*) or an embodiment combining software and hardware aspects that may generally be referred to herein as a "system." Furthermore, embodiments of the present invention may take the form of a computer program product on a computer-readable medium having computer-usable program code embodied in the medium.

[0048] Any suitable computer-readable medium may be utilized. The computer-readable medium may be, for example but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific examples of the computer readable medium include, but are not limited to, the following: an electrical connection having one or more wires; a tangible storage medium such as a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), a compact disc read-only memory (CD-ROM), or other optical or magnetic storage device; or transmission media such as those supporting the Internet or an intranet. Note that the computer-readable medium could even be paper or another suitable medium upon which the program is printed, as the program can be electronically captured, via, for instance, optical scanning of the paper or other medium, then compiled, interpreted, or otherwise processed in a suitable manner, if necessary, and then stored in a computer memory.

[0049] Computer program code for carrying out operations of embodiments of the present invention may be written in an object oriented, scripted or unscripted programming language such as Java, Perl, Smalltalk, C++, or the like. However, the computer program code for carrying out operations of embodiments of the present invention may also be written in

conventional procedural programming languages, such as the "C" programming language or similar programming languages.

[0050] Embodiments of the present invention are described below with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems), and computer program products. It may be understood that each block of the flowchart illustrations and/or block diagrams, and/or combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create mechanisms for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0051] These computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer readable memory produce an article of manufacture including instruction means which implement the function/act specified in the flowchart and/or block diagram block(s).

[0052] The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions which execute on the computer or other programmable apparatus provide steps for implementing the functions/acts specified in the flowchart and/or block diagram block(s). Alternatively, computer program implemented steps or acts may be combined with operator or human implemented steps or acts in order to carry out an embodiment of the invention.

[0053] Thus, apparatus, systems, methods and computer program products are herein disclosed that provide for determining suggestive potential custodians for a case/matter within a comprehensive electronic discovery system. The determination is based on a comparison of data, such as personnel data or like, associated with individuals already added to the case and all other individuals in the enterprise, who form the group of potential custodians. In addition, the comparison of data occurs over the specified timeframe defined by the case to insure that the

suggested potential custodians are within the confines to the timeframe of the case. Once the suggested potential custodians have been determined they are presented to the electronic associate charged with added potential custodians to the case and the associate can make a final determination as to whether to add or not add the suggested potential custodian to the case.

[0054] Additional embodiments of the invention provide for determining preservation notice recipients based on a comparison of data, such as personnel data or the like, associated with individuals that have already been identified to receive the preservation notice and other custodians and/or potential custodians. In addition, the determination takes into account the preservation notice history of the custodian and/or potential custodian being considered as a recipient of a specific preservation notice.

[0055] **FIG. 1** provides a schematic diagram of an apparatus 10 configured for determining suggested potential custodians for a case or matter in an enterprise-wide electronic discovery system; in accordance with embodiments of the present invention. As used herein, the term "custodian" refers to an individual who has access to data on a data storage device. In addition to having access, it will be understood that a custodian usually reads, writes, creates, stores, deletes, alters, changes, processes, and/or otherwise manipulates data on the data storage device. An individual is a "potential custodian" if they have been identified as someone of interest to a case or matter. Once a "potential custodian" has been identified, they are added to the respective case/matter. The "potential custodian" is elevated to an actual "custodian" once a decision is made to collect data from that particular individual. As used herein, the term "enterprise" refers to a business, organization, association, and/or some other entity that usually includes a plurality of custodians. For example, in some embodiments, an enterprise is a major corporation that employs several thousands of employees, some or all of which may be data custodians. Unless explicitly stated otherwise, it will also be understood that the term "custodian" usually refers to an employee (current, former, full-time, part-time, contract, etc.) of an enterprise.

[0056] The apparatus 10, which may comprise more than one computing devices, includes a computing platform 12 having at least one processor 14 and a memory 16. In specific embodiments of the invention the apparatus may take the form of an electronic discovery manager server 110 as shown in and described in relation to **FIG. 4**. The memory 16 of apparatus 10 stores electronic discovery manager application 20 which is configured to receive

information from electronic discovery associates as a means of defining cases 22 and/or matters within the electronic discovery system. As such, electronic discovery manager application 20 is configured to receive first inputs that define a case 22. Included within the definition of the case 22 is the timeframe 24 over which the case occurs. The case timeframe 24 may be defined by specific dates, such as a case opening date and/or a case closing date or the timeframe may be loosely defined, such as open ended on either the opening or closing end. The electronic discovery manager application 20 is additionally configured to receive second inputs that add one or more potential custodians 26 to the case 22. The initial potential custodians 26 added a case 22 are typically, but not required to be, individuals readily identified and closely related to the case.

[0057] In accordance with specific embodiments of the invention, the case and the custodian data received by the electronic discovery manage application are stored in case database 42 and custodian database 46 within the electronic discovery system database 40. The case database 42 includes a plurality of case files 44 associated with each case in the system and each case file has one or more custodians 48 assigned to the case. The custodian database 46 includes a plurality of custodian files 48 and each custodian file has one or more cases 49 associated with the custodian. As noted, the case database 42 is linked to the custodian database 46, such that when a custodian is added to a case, the addition of the custodian is reflected in the respective case file 44 of the case database 42 and the case is reflected in the respective custodian file 48 of the custodian database 46. The link between the case database 42 and the custodian database 46 provides for leveraging future cases/matters based on previous historical case/matter data. The concept of linking the case database 42 to the custodian database 46 is discussed further in relation to **FIGS. 4 - 12.**

[0058] The memory 16 of apparatus 10 additionally includes suggested potential custodian application 30, which includes suggested potential custodian routine 32. The suggested potential custodian routine 32 is configured to determine a plurality of suggested potential custodians 36 based on association 34 with previously added potential custodians 26 and the timeframe 24 associated with the case. The association 34 between previously added potential custodians 26 and suggested potential custodians 36 may include, but is not limited to, employment chain-of-command, *e.g.*, same superiors or subordinates between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24; employment

location, *e.g.*, same city, site, building, floor, *etc.*, between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24; employment classification, *e.g.*, same job code, job title etc. between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24; same data source accessibility between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24 and/or similar data content generated by previously added custodians 26 and suggested potential custodians 36 during case timeframe 24. The suggested potential custodian routine 32 access enterprise database 38, which may be any database throughout the enterprise including personnel databases and the like, to provide for an association between a previously added custodian 26 and the suggested potential custodians 36.

[0059] Further suggested potential custodian application 30 is configured to provide an output for presentation of the suggested potential custodians 36. In one specific embodiment the output for presentation of the suggested potential custodians 36 is a display, such as a pop-up window or the like, that is presented in conjunction with a user, such as an electronic discovery associate, adding potential custodians 26 to the case 22. In this regard, the suggested potential custodians are presented to the electronic discovery associate during the initial phase of adding potential custodians to the case and the list of suggested potential custodians may be refined (*i.e.*, expanded or limited), in real-time, as more potential custodians 26 are added to the case 22.

[0060] Referring to **FIG. 2**, shown is a more detailed block diagram of apparatus 10, specifically electronic discovery manager server 110 embodied within an electronic discovery enterprise system 100, according to embodiments of the present invention. The electronic discovery manager server 110 is configured to determine suggested potential custodians to add to a case/matter within the electronic discovery system 100 and, in optional embodiments, determine suggested preservation notice recipients. In addition to providing greater detail, **FIG. 2** highlights various alternate embodiments. The electronic discovery manager server 110 may include one or more of any type of computerized device. The present apparatus and methods can accordingly be performed on any form of computing device.

[0061] The server 110 includes computing platform 12 that can receive and execute routines and applications. Computing platform 12 includes memory 16, which may comprise volatile and non-volatile memory, such as read-only and/or random-access memory (RAM and ROM), EPROM, EEPROM, flash cards, or any memory common to computer platforms.

Further, memory 16 may include one or more flash memory cells, or may be any secondary or tertiary storage device, such as magnetic media, optical media, tape, or soft or hard disk.

[0062] Further, computing platform 12 also includes processor 14, which may be an application-specific integrated circuit ("ASIC"), or other chipset, processor, logic circuit, or other data processing device. Processor 14 or other processor such as ASIC may execute an application programming interface ("API") 50 that interfaces with any resident programs, such as electronic discovery manager application 20, suggested potential custodian application 30, suggested preservation notice recipient application 70 or the like stored in the memory 16 of the electronic discovery manager server 110.

[0063] Processor 14 includes various processing subsystems 52 embodied in hardware, firmware, software, and combinations thereof, that enable the functionality of server 110 and the operability of the server on a network. For example, processing subsystems 42 allow for initiating and maintaining communications and exchanging data with other networked devices. For the disclosed aspects, processing subsystems 52 of processor 14 may include any subsystem used in conjunction with electronic discovery manager application 20, suggested potential custodian application 30, suggested preservation notice recipient application 70 or subcomponents or sub-modules thereof.

[0064] Computer platform 12 additionally includes communications module 54 embodied in hardware, firmware, software, and combinations thereof, that enables communications among the various components of the electronic discovery manager server 110, as well as between the other devices in the electronic discovery system 100. Thus, communication module 54 may include the requisite hardware, firmware, software and/or combinations thereof for establishing a network communication connection.

[0065] As previously noted, the memory 16 of computing platform 12 stores electronic discovery manager application 20 which is configured to receive information from users, such as electronic discovery associates or the like, that define cases 22 and/or matters within the electronic discovery system. As such, electronic discovery manager application 20 is configured to receive first inputs that define a case 22. Included within the definition of the case 22 is the timeframe 24 over which the case occurs. The electronic discovery manager application 20 is additionally configured to receive second inputs that add one or more potential custodians 26 to

the case 22. The initial potential custodians 26 added a case 22 are typically, but not required to be, individuals readily identified and closely related to the case.

[0066] The memory 16 of apparatus 10 additionally includes suggested potential custodian application 30, which includes suggested potential custodian routine 32. The suggested potential custodian routine 32 is configured to determine a plurality of suggested potential custodians 36 based on one or more associations 34 with previously added potential custodians 26 and the timeframe 24 associated with the case. The association 34 may comprise a determined data match 60 between predetermined data associated with previously added potential custodians 26 and suggested potential custodians 36.

[0067] The predetermined data may include, but is not limited to, employment chain-of-command, *e.g.*, same superiors or subordinates between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24; employment location, *e.g.*, same city, site, building, floor, *etc.*, between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24; employment classification, *e.g.*, same job code, job title etc. between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24; same data source accessibility between previously added custodians 26 and suggested potential custodians 36 during case timeframe 24 and/or similar data content generated by previously added custodians 26 and suggested potential custodians 36 during case timeframe 24. The suggested potential custodian routine 32 accesses the enterprise database 38, which may be any database throughout the enterprise including personnel databases and the like, to provide for an association between a previously added custodian 26 and the suggested potential custodians 36.

[0068] In specific embodiments of the invention a data match threshold 62 may be defined as a means of determining whether an association exists between previously added potential custodians 26 and suggested potential custodians 34. For example, in the predetermined data includes four categories/criteria the data match threshold for an association to be determined may be a match in three of the four categories. In a specific example, if the four categories/criteria are (1) employment chain-of-command/supervisor, (2) employment location; (3) employment job title; and (4) data source accessibility, an association between a previously added potential custodian 26 and a suggested potential custodian 34 may be determined if the data match threshold 62 is met or exceeded. If the data match threshold is three, less than three

matches results in no association between the previous potential custodian 26 and the suggested potential custodian 34 (in other words, the individual of note does not rise to the level of suggested potential custodian 34) and three or more matches results in an association between the previous potential custodian 26 and the suggested potential custodian 34.

[0069] In further specific embodiments, a data match relevancy weight 64 may be applied to each predetermined data category/criteria and the sum of the weighted data match categories/criteria results in a suggested potential custodian score 66. For example, employment chain-of-command/supervisor may have a weight of ten; employment location may have a weight of two and all other data match categories/criteria may have a weight of one. If it is determined that the previously added potential custodian 26 and the suggested potential custodian 34 had the same employment chain-of command/supervisor and the same employment location but shared no other data match categories/criteria, the suggested potential custodian score 66 would be twelve (*i.e.*, the sum of ten and two).

[0070] In specific embodiments of the invention the predetermined data may include sub-categories/sub-criteria that provide for predetermined data matches 60. For example, the previously added potential custodian 26 may not have to be located in the same building as the suggested potential custodian 36 but may however be located at the same site within a different building during the timeframe of the case. Thus, in certain embodiments, sub-category matches may be used to determine an association between previously added potential custodians 26 and suggested potential custodians 34. For example, if the embodiment implements data match relevancy weighting 64, the weight assigned to the category/criteria or a portion less than the entire weight assigned to the category/criteria may be assigned based on the sub-category match between the previously added potential custodian 26 and suggested potential custodian 34.

[0071] In other embodiments of the invention, the data match relevancy weight 64 can be determined and assigned to a data category/criteria and or data sub-category/sub-category on-the-fly, based on how similar each data element is within a data category or data sub-category for the custodians that been previously added to the case.

[0072] As previously noted, embodiments of the present invention further provide for the output of suggested potential custodians 34 to be refined, in real-time or the like, based on the addition of one or more potential custodians 26 to the case/matter 22. In this regard, the suggested potential custodian routine may include a clustering algorithm or another like

algorithm that provides for suggested potential custodians to be determined based on their association with more than one previously added potential custodians 26. Examples of other algorithms include, but are not limited to, Bayesian algorithms, logistic algorithms, decision tree algorithms, regression algorithms and the like. The clustering algorithm provides for assessing associations between all of the previously added potential custodians 26 and the database of suggested potential custodians (*i.e.*, all other employees of the enterprise) to determine association with all the previously added potential custodians 26. Thus, if an individual being considered as a suggested potential custodian has stronger associations with only a few of the previously added potential custodians or weaker associations with many of the previously added potential custodians, the clustering aspect of the algorithm may give rise to identifying the individual as a suggested potential custodian. Once suggested the user, such as an electronic discovery associate, can choose to add the potential custodian to the case/matter. In this regard, as more potential custodians are added to case which have little or no association with a suggested potential custodian 36, the suggested potential custodian may fall off or otherwise be removed from the listing of suggested potential custodians 36 unless the association with the previously added potential custodians is strong. A strong association may be defined as having met a predetermined data match threshold 62 and/or having a predetermined suggested potential custodian score 66 or the like.

[0073] In specific embodiments of the method more than one method may be implemented to determine suggested potential custodians 36. For example, multiple different algorithms may be implemented to determine suggested potential custodians and the results presented by listings grouped by algorithm or the results of each algorithm may be combined, such as by averaging, weighted averaging or the like.

[0074] Further, as previously noted, suggested potential custodian application 30 is configured to provide an output for presentation of the suggested potential custodians 36. In one specific embodiment the output for presentation of the suggested potential custodians 36 is a display, such as a pop-up window or the like, that is presented in conjunction with a user, such as an electronic discovery associate, adding potential custodians 26 to the case 22. In this regard, the suggested potential custodians are presented to the electronic discovery associate during the initial phase of adding potential custodians to the case and the list of suggested potential custodians may be refined and/or expanded, in real-time, as more potential custodians 26 are

added to the case 22. In such embodiments, in which suggested potential custodian scoring 66 is implemented, the output of the determined plurality of suggested potential custodians may include a ranked listing of suggested potential custodians with the highest scoring potential custodians appearing at the top of the list. A threshold score may be assigned such that only suggested potential custodian 34 having a suggested potential custodian score 66 at or above a minimal threshold are presented in the output.

[0075] In addition, in those embodiments in which the output of suggested potential custodians 34 are refined, in real-time or the like, based on the addition of one or more potential custodians 26 to the case/matter 22 and suggested potential custodian scores 66 are implemented, the scores 66 are based on associations 32 and/or predetermined data matches 60 with all of the previously added potential custodians 26. As such the suggested potential custodian score 66 is refined, in real-time, as new potential custodians 26 are added to the case/matter 22. Thus, the ranked order of the listing of suggested potential custodians 26, in which the ranking is based on suggested potential custodian scores 66, may dynamically change over time as new potential custodians 26 are added to the case 22.

[0076] Additionally, according to specific embodiments, memory 16 of electronic discovery manager server 110 includes suggested preservation notice recipient application 70 that is configured to determine suggested preservation notice recipients 72 from amongst potential custodians 26 previously added to the case based on an association with one or more previously identified recipients of the preservation notice and previous preservation notice history of the one or more suggested potential recipients. In this regard, the suggested preservation notice recipient application 70 takes into account associations with previous recipients and which types of preservation notices have been received by a potential custodian considered for recipient in the case/matter at hand and/or in previous cases/matters.

[0077] In such embodiments, the preservation notice receipt application may be configured to determine suggested preservation notice recipients in any of the manners previously discussed in relation to determining suggested potential custodians. For example, suggested preservation notice recipients may be determined based on a match between predetermined data associated with at least one previously identified recipients and the potential custodian being considered as a suggested recipient. Moreover, a predefined threshold of matches may be set, such that the predefined threshold of matches need to be determined to exist

between the predetermined data associated with the previously identified recipients and the potential custodians being considered as suggested recipients in order for determination of a suggested preservation notice recipient. Similar, to the determination of suggested potential custodians, the predetermined data associated with the previous identified recipients and the potential custodians considered as recipients may include, but is not limited to, employment chain-of-command, employment locale, employment classification, data source accessibility, generated data content and the like. In still further embodiments, the preservation notice recipient application 70 may incorporate relevancy weighting to determine suggested preservation notice recipients and subsequent suggested preservation notice recipient scoring for the purpose of listing suggested preservation notice recipients in a ranked order.

[0078] Referring to **FIG. 3** a flow diagram is depicted of a method 80 for determining suggested potential custodians for a case/matter in an electronic discovery system, in accordance with embodiments of the present invention. At Event 82, first inputs are received that define a case within the electronic discovery system and include a timeframe associated with the case. The timeframe may include a specific time range based on specific dates or the timeframe may be open-ended in terms of a start date or a close date.

[0079] At Event 84, second inputs are received that add one or more potential custodians to the case. As previously noted the initial custodians added to a case are typically individuals within the enterprise highly relevant to the case. Additionally, potential custodians added to a case become actual custodians once a decision is made to acquire data from the individual or from devices associated with the individual.

[0080] At Event 86, a plurality of suggested potential custodians is determined for the case. The determination of the suggested potential custodians is based on an association with the previously added custodians and the timeframe associated with the case. The association may be a match between predetermined data associated with the previously added potential custodians and the individuals being considered as suggested potential custodians. Moreover, the determination of suggested potential custodians may further include matching a predetermined threshold of the predetermined data. The predetermined data may include, but is not limited to, employment chain-of-command, employment location, employment classification, data source accessibility, generated data content and the like.

[0081] Additionally, determination of the suggested potential custodians may include weighting the predetermined data based on respective relevancy to association. In such embodiments, the weighted values of the determined data may be summed to result in a suggested potential custodian score. A threshold score may be set, such that, a suggested potential custodian is determined if the threshold score is met or exceeded.

[0082] In further embodiments clustering techniques may be implemented to determining associations between a plurality of previously added potential custodians and individual being considered for suggested potential custodians. Clustering techniques allow for identification of associations between the plurality of previously added potential custodians and the individual being considered for suggested potential custodians, as well as, identification of the strength of the associations.

[0083] At Event 88, an output is provided of the determined plurality of suggested potential custodians. In accordance with specific embodiments, the output may be a visual presentation, such as a pop-up window or the like, displayed in conjunction with a display that provides for the addition of potential custodians to the case. In instances in which suggested potential custodian scoring is implemented, the visual presentation may provide for a ranked listing of the suggested potential custodians based on their respective suggested potential custodian score. Additionally, the listing of suggested potential custodians may be refined and/or expanded in real-time based on the addition of potential custodians to the case.

[0084] At optional Event 90, third inputs are received that add one or more of the suggested potential custodians to the case. The third inputs may be at the discretion of the user, for example, an electronic discovery associate or the third inputs may be automated inputs based on predetermined criteria for adding a suggested potential custodian to the case/matter.

[0085] In addition, once one or more suggested custodians has been added, the method returns to Event 86 for further determination of suggested potential custodians using the new information associated with the latest custodian(s) added. In this regard, the determination of suggested potential custodians occurs in real-time as custodians are added to the case and, thus, the suggested potential custodians that are determined become more closely associated with the group of added custodians based on determination taking into account the latest custodian(s) added to the case.

[0086] As a means of providing an overview of a composite, enterprise-wide electronic discovery system, which may be implemented in conjunction with the computer monitoring of network status described above, **FIGS. 4 - 12** are herein provide and described in detail. It should be noted that the system shown and described in relation to **FIGS. 4 - 12** are by way of example only and, as such, presently described embodiments of the invention may be embodied within other electronic discovery systems or only implement select portions of the electronic discovery system herein described.

[0087] **FIG. 4** illustrates an exemplary electronic discovery system 100 in accordance with an embodiment of the invention. In some embodiments, the environment of the electronic discovery system 100 is the information technology platform of an enterprise, for example a national or multi-national corporation, and includes a multitude of servers, machines, and network storage devices in communication with one another over a communication network. In particular, an electronic discovery management server 110, at least one database server 120, a collections server 130, enterprise personal computers 140, enterprise file servers 150, including at least one personal network storage area and at least one shared network storage area, enterprise email servers 160, a conversion services server 170, a short-term staging drive 180, and a long-term network storage network 190 are all in communication over a communication network 102. The communication network 102 may be a wide area network, including the Internet, a local area network or intranet, a wireless network, or the like.

[0088] As shown in the block diagram of **FIG. 5**, the electronic discovery management server 110 provides user interface management for via user interface 118. In some embodiments, the electronic discovery management server 110 is a web server that can be accessed via a web browser. In one particular embodiment, the electronic discovery management server 110 is an intranet website server that may be accessed utilizing a web browser on a machine within the enterprise. Through the electronic discovery management server 110, the user interface 118 may be presented to a user for the purposes of managing the electronic discovery process and all processes described herein that are inherent thereto. For illustrative purposes, it may be assumed herein that the primary user interacting with the user interface 118 is an employee or contractor of the company who serves an electronic discovery management role, and hereafter is referred to as the "e-discovery manager." As discussed in greater detail below, the e-discovery manager may utilize the user interface 118 to manage cases,

custodians, collections, and collected data. It should be appreciated, however, that any individual could use the user interface 118 to perform the manual functions herein attributed to the e-discovery manager, and, indeed, that an automated process could perform those functions as well.

[0089] Referring again to **FIG. 4**, the electronic discovery management server 110 is in communication with the database server 120 and the collections server 130 via the communication network 102. The database server 120, as shown in the block diagram of **FIG. 6**, is configured to provide database services for the electronic discovery management server 110, including housing the Unified Directory/custodian database 122, which includes data relating to individual custodians, the case database 124, which includes data relating to particular cases, and ongoing collections database 126, which includes data relating to collections being undertaken by the collections server 130. Each of the foregoing databases within the database server 120 is discussed in detail below. It should be understood that multiple database servers could be employed instead of a single database server, and reference to a single database server is for illustrative and convenience purposes only. For example, the Unified Directory 122 could be stored in one database server and the ongoing collections data 126 could be stored in another database server.

[0090] Regardless of the number of database servers employed, it is an object of embodiments of the present invention that data relating to custodians and cases be stored in the database server 120 independently. While custodian data in the Unified Directory 122 and case data in the case database 124 may be linked or correlated within the database server 120, for example, when custodians are assigned to particular cases, custodians may be managed separately from cases. Therefore, when a case is initialized and a custodian is assigned to the case, information for that custodian (such as data storage locations for that custodian) is accessed by the electronic discovery management server 110 in the Unified Directory 122 in the database server 120 and linked to the particular case, rather than manually input by the e-discovery manager into the case.

[0091] Furthermore, in addition to separating (but allowing linkage of) custodian management and case management processes, and as discussed further below, data management processes relating to the collection of data from custodian storage locations during electronic discovery are also separated from case management and custodian management processes. In

this regard, the data collected from a particular custodian is stored separately from both the custodian information and any relevant case information (as discussed below, it is stored in long-term network storage network 190), but is linked to a custodian, which is in turn linked to one or more cases. This is advantageous because in the event a particular custodian is assigned to multiple cases, data collected from the custodian may be shared with the other case(s) to which the custodian is assigned. Therefore, the various processes and components of the electronic discovery system 100 may be categorized within one of case management, custodian management, or data management. And even though cases, custodians, and collected data may all be managed separately, there are necessarily links between the various datastores to allow management of the overall electronic discovery process.

[0092] CUSTODIAN

[0093] With regard to custodian management, according to some embodiments of the present invention, the Unified Directory/custodian database 122 houses information relating to all potential custodians within the enterprise and the locations where those custodians store data. The information stored in the Unified Directory 122 may include for a particular custodian, for example, the custodian's name, position, human resources identifier (a unique number for each employee of the enterprise), employment location, domain, email addresses, network user identification, personal computer(s) name, paths of network storage devices used by the custodian, including Shared Drives and HomeSpaces, work history, related persons (such as managers, team members or subordinates), and any other information that may be relevant to the discovery process. Since the human resources identifier is always unique for each custodian, in some embodiments, the Unified Directory 122 may be organized around the human resources identifier. All of the information relating to how the Unified Directory 122 is generated is a multi-step process that utilizes multiple applications and methods of identifying relevant information.

[0094] For example, the electronic discovery management server 110 or the database server 120 may interface with the computer databases of the human resources computer systems of the enterprise to copy the information from the human resources databases into the Unified Directory 122. In some embodiments, the electronic discovery management server 110 may also reach out to a network directory, such as Windows Active Directory, to identify network resources related to particular custodians and integrate this information into the custodian entries

including the copied human resources information. Information for the Unified Directory 122 may also be obtained from the managers of the information technology network, *i.e.*, those individuals responsible for setting up email accounts for custodians and managing the various file servers of the enterprise. Furthermore, in addition to retrieving information in the manners described above, in some embodiments, information in the Unified Directory 122 is generated through applications initialized and/or deployed by the electronic discovery management server 110. In particular, in some embodiments, as shown in **FIG. 4**, a profile scanning application 112, and a mapping application 114 are provided.

[0095] The profile scanning application 112 may be deployed by the electronic discovery management server 110 and is configured to crawl the communication network 102, scan each of the enterprise personal computers 140, and transmit to the database server 120 identifying information about each computer, such as computer name and IP address, and a list of all profiles, including demographics information, (or network user identification) associated with each computer. According to different embodiments, the profile scanning application 112 may be run on the electronic discovery management server 110, the collection server 130, or another server in the communication network 102. In some embodiments, the profile scanning application 112 is further configured to identify and transmit to the database server 120 the most recent date and time at which a particular profile was logged on to the machine. When information relating to a particular computer is received by the database server 120, the database server 120 uses the profile information, which may include several user identifications, to link the particular computer to the custodians in the Unified Directory 122 associated with those user identifications. The database server 120 may also record in each custodian's entry in the Unified Directory 122 the last time the computer was accessed by the custodian, according to the profile information transmitted by the profile scanning application 112. Thus, the profile scanning application 112 ultimately generates a list of personal computers used by each custodian, and this list may be presented to the e-discovery manager when a collection of a custodian's local machine(s) is initialized, as discussed in detail below.

[0096] In accordance with some embodiments of the invention, the mapping application 114 is configured to crawl the communication network 102 and examine the enterprise file servers 150 residing on the communication network 102 to locate and identify the path of any personal network storage area on each server. As used herein, a personal network storage area is

a network storage area associated with a single user who reads data from or writes data to it. Personal network storage areas may be in the form of network storage devices or folders or other resources within a network storage device and may be referred to hereafter for clarity purposes as "HomeSpaces." According to different embodiments, the mapping application 114 may be run on the electronic discovery management server 110, the collection server 130, or another server in the communication network 102. In some embodiments, the mapping application 114 is a Windows service that is scheduled to execute through use of Windows Scheduled Task. As the mapping application 114 crawls the communication network 102, it is configured to examine each file server and transmit to the database server 120 the path of any network storage area within the plurality of servers 134 that it positively identifies as a HomeSpace. In some embodiments, the mapping application 114 is configured to explore the enterprise file servers 150 by obtaining and reviewing the directories on each server and evaluating the paths of each network storage area therein, including folders and other storage devices and resources.

[0097] With regard to identifying a particular network storage area as a HomeSpace, according to some embodiments, the mapping application 114 is configured to utilize conventional naming techniques for paths in the communication network 102 to identify those paths of network storage areas within the enterprise file servers 150 that include an indicator, based on the conventional naming techniques, that the particular storage areas associated with those paths are accessed and used by only one user, and are therefore HomeSpaces. In accordance with some embodiments of the invention, each user of the communication network 102 is assigned to at least one user identification and those user identifications are the indicators that the mapping application 114 attempts to locate within paths when identifying HomeSpaces. In such embodiments, it is the convention that the paths of HomeSpaces on the communication network 102 include the user's user identification. On the other hand, paths of shared network storage areas do not include user identifications. Therefore, the mapping application 114 may explore the directories of each server within the plurality of servers, evaluate each path in turn, and make a determination as to whether or not the path includes a user identification.

[0098] If it is determined that the path includes the designated indicator, for example, a user identification, the mapping application 114 is configured to positively identify the particular network storage area identified by that path as a HomeSpace and transmit to the database server 120 the particular user identification and the path of the HomeSpace. When that information is

received by the database server 120, the database server 120 uses the user identification to link the particular HomeSpace to the custodian in the Unified Directory 122 associated with that user identification. In some embodiments, the mapping application 114 is also configured to recognize and transmit, and the database server 120 is configured to house, an indication of the last time the HomeSpace was accessed by the particular user, for example, the last time any data was read from and/or written to the HomeSpace. Additionally, in some embodiments, the mapping application 114 is configured to recognize when multiple paths map to the same network storage area. The collection server 130 compares paths for the same user to determine if duplicative entries exist. This advantageously enables avoidance of multiple collections of the same data. Thus, the profile scanning application 112 ultimately generates a list of HomeSpaces used by each custodian, and this list may be presented to the e-discovery manager when a collection of a custodian's HomeSpaces is initialized, as discussed in detail below.

[0099] In addition to storing a list of personal computers and HomeSpaces used by a particular custodian, which lists were generated by the profile scanning application 112 and the mapping application 114 respectively, in accordance with some embodiments of the present invention, the database server 120 is also configured to store a list of any shared network storage areas used by the custodian. As used herein, a shared network storage area is a network storage area associated with multiple users who read data from and/or write data to it. Shared network storage areas may also be in the form of network storage devices or folders or other resources within network storage devices and may be referred to hereafter for clarity purposes as "Shared Drives." The user interface 118 is configured to receive a path of a Shared Drive input by the e-discovery manager and store the path in the Unified Directory 122 in relation to one or more custodians' human resources identifier(s). More particularly, in some embodiments, once a particular user of the communication network 102 is chosen for the collection process, the e-discovery manager may undertake to identify the particular shared network resources that that individual is using, and eventually, the paths associated with those shared network resources. This may be accomplished through conversations with the particular individual, by utilizing data returned from the local collection application 132 executed on collection server 130 (shown in the block diagram of **FIG. 7**) deployed to the particular user's machine (as discussed in detail below), and/or by utilizing a file browsing application 116 executed on electronic discovery manager server 110 (as shown in **FIG. 5**).

[00100] According to some embodiments of the present invention, the file browsing application 116 is configured to be utilized by the e-discovery manager through the user interface 118. The file browsing application 116 gives the e-discovery manager elevated authority within the communication network 102 to access, in a limited manner, the enterprise file servers 150 within the communication network 102. While the file browsing application 116 may not allow access to the actual files stored on certain file servers, it allows the e-discovery manager to browse through the directories of the file servers 150, locate files that have been accessed by the custodian, and determine the size of the files. In accordance with some embodiments, the e-discovery manager may initially have a general idea of a particular file server within the enterprise file servers 150 that the custodian has used in the past. For example, the custodian may communicate to the e-discovery manager a particular folder name and/or drive name on which he/she has stored files. Additionally, in some embodiments, the e-discovery manager may have already undertaken a local collection process on the custodian's machine, wherein the local collection application 132 returned a list of the network resources that the user of that machine has used. In that event, the e-discovery manager may be aware of the particular drive referenced by the user. The e-discovery manager may then employ the file browsing application 116 to browse out to the particular drive mentioned, scan the folders for any folder having a name resembling that name given by the user, identify any particular files created by and/or accessed by the user, determine the size of such files, and retrieve the path of any folder (or Shared Drive) including data belonging to the user.

[00101] The retrieved paths of the Shared Drives may then be added, either manually or automatically, to the Unified Directory 122 in the database server 120. Thus, the Unified Directory 122 may store in connection with one custodian (and in particular in relation to the custodian's human resources identifier) a list of the personal computers, HomeSpaces, and Shared Drives associated with that custodian. Each of these locations is a potential source of data stored by the custodian, and once an investigation or collection of a custodian is initiated, the location information stored in the Unified Directory 122 may be accessed to determine the particular storage locations that need to be addressed during the investigation/collection. This is advantageous as it allows a completely automated investigation/collection process, rather than relying on the e-discovery manager to manually input the targeted machines and file servers at the time of collection.

[00102] It should be noted that the Unified Directory 122 may be regularly or continuously updated as new information is gathered using the applications described herein. More particularly, the electronic discovery management server 110 may be configured to automatically retrieve data from the human resources databases and Active Directory and any other relevant sources, such as information technology directories or lists, as well as deploy the profile scanning application 112 and the mapping application 114, at regularly scheduled intervals. Alternatively, rather than periodically retrieving data from the various data sources such as the human resources databases, the system 100 may be configured such that the database server 120 is continuously interfacing with the data sources such that the Unified Directory 122 is updated in real-time as the data within the data sources change. In either instance, each of the feeds of information into the Unified Directory 122 is regularly updated to ensure that the data in the Unified Directory 122 is current.

[00103] In some embodiments, the database server 120 is configured such that all historical data relating to a custodian is stored in relation to that custodian's human resources identifier in the Unified Directory 122. Thus, when the feeds of information into the Unified Directory 122 are updated, in the event data relating to the custodian has changed, the database server 120 is configured to store in the Unified Directory 122 the new data and any relevant metadata, including, for example, the time and date of the change, as well as maintain a record of the old data so that it is still a part of the custodian's profile in the Unified Directory 122. For example, in the event the profile scanning application 114 identifies a new personal computer associated with a custodian and one of the personal computers associated with the custodian previously is no longer identified, the database server 120 is configured to store in the Unified Directory 122 the information for each computer, as well as indications as to when the new computer was first identified and when the old computer was no longer identified. In this way, the custodian profile within the Unified Database 122 may include a history of the personal computers used by the custodian. Such information may be relevant at the time of investigation or collection of the custodian.

[00104] One feed of information into the Unified Directory 122 which is particularly relevant to electronic discovery is employment status. According to some embodiments, when the feed of information from the human resources databases to the Unified Directory 122 includes a change as to employment status of a particular custodian, the electronic discovery

management server 110 is configured to recognize the change and possibly perform particular functions in response. More specifically, in the event it is recorded in the Unified Directory 122 that the employment status of a particular custodian changes from active to terminated, the electronic discovery management server 110 is configured to determine whether the custodian is assigned to any case or matter, and, if so, to transmit to the designated manager or contact for the case or matter an electronic communication notifying the manager of the terminated status and inquiring as to whether the manager would like the terminated custodian's data collected. In the event the manager responds in the affirmative, the electronic discovery management server 110 is configured to automatically initiate the various collection processes of the present invention. Therefore, the custodian's data may be advantageously collected prior to any destruction or unavailability that could be caused by the termination. Alternatively, in other embodiments, the electronic discovery management server 110 may not communicate with the manager and may automatically initiate collection upon recognizing a change in employment status.

[00105] CASE

[00106] With regard to case management processes, according to some embodiments, a case may be initialized by the e-discovery manager utilizing the user interface 118. In this regard, the e-discovery manager may enter into the user interface 118 certain information about a particular matter or case, such as a case name and/or number, a short description of the matter/case, a legal identifier, the particular requester (*i.e.*, who asked for the case to be opened), managers or contacts for the matter (*i.e.*, individuals involved in the substance of the matter rather than the process, like the e-discovery manager), custodians, *etc.* The electronic discovery management server 110 is configured to store this information in the case database 124 in the database server 120. The case database 124 is configured to house this information such that all information relating to a particular matter or case is related within the case database 124 and a user can use the user interface 118 to view a profile of the matter or case including all the information.

[00107] Once the matter and/or case has been initialized, the e-discovery manager may add custodians to the matter or case. In some embodiments, the electronic discovery management server 110 is configured to add numerous custodians to a single matter or case at one time. In this regard, the e-discovery manager may use the user interface 118 to enter in identifying information about the custodians. The identifying information for each custodian

does not have to be of the same type. For example, a name may be entered for one custodian, an email address for another, a network user identification for another, and a human resources identifier for another. The user interface 118 is configured to receive the identifying information in different input areas depending upon the type of identifying information being received. The electronic discovery management server 110 is configured to use the input information to search the Unified Directory 122 in the database server 120 to determine which custodians are associated with the input information. In the case of a human resources identifier being entered, only one custodian in the Unified Directory 122 may be a match. On the other hand, in the case of a name being entered, multiple custodians may be a match.

[00108] The electronic discovery management server 110, after searching the Unified Directory 122 with the input identifying information, is configured to present through the user interface 118 a list of all custodians matching the input identifying information. In the event only one match was returned for a particular set of input identifying information, the electronic discovery management server 110 is configured to automatically select the custodian to be added to the case or matter. On the other hand, in the event more than one match was located for a particular set of input identifying information, then the multiple matches may be presented together to the e-discovery manager through the user interface 118 and marked so that the e-discovery manager must review the multiple custodian profiles associated with the matches to determine the correct custodian that should be added to the case or matter. In doing so, the e-discovery manager may consider the other information in the profiles, such as corporate title, work location, associated custodians, *etc.* Such information can inform the e-discovery manager as to whether the located custodian is the one intended. The e-discovery manager may then select the correct custodian for addition to the case or matter and confirm that all custodians selected may be added to the case or matter. According to some embodiments, "adding" a custodian to a case or matter involves linking correlating the custodian profile in the Unified Directory 122 to the case or matter in the Case database 124.

[00109] According to some embodiments, upon adding custodians to a matter, the electronic discovery management server 110 is configured to initiate the transmission of preservation notices and surveys to the custodians. In this regard, preservation notices and surveys relevant to the particular case or matter are stored in or linked to the case profile in the case database 124. Transmission of the preservation notices and surveys to custodians added to

the case may be automated, for example, there may be preset instructions within the case profile that cause the electronic discovery management server 110 to transmit a particular preservation notice and survey at a particular date or time or upon a particular triggering event, such as a custodian being added to the case, or the e-discovery manager may manually cause the preservation notices and surveys to be transmitted. In some embodiments, the electronic discovery management server 110 is configured to transmit the preservation notices and surveys via a standard email function. The surveys may be tied to the preservation notices such that they are transmitted to custodians together, and one survey may be tied to more than one preservation notice. When a custodian responds to a survey, the survey response is received by the electronic discovery management server 110 and stored in relation to the relevant custodian in the case profile in the case database 124. Furthermore, the electronic discovery management server 110 may be configured to store all or a portion of the data received in the survey response in the Unified Directory 122 in the custodian's profile.

[00110] According to some embodiments, each transmission of a preservation notice and survey to a custodian, and each corresponding response, is tracked in the relevant case profile in the case database 124. The electronic discovery management server 110 may also be configured to transmit reminder notices if responses to the surveys are not received within a predefined period of time. The electronic discovery management server 110 may also be configured to schedule reminder notices to be sent to custodians to periodically refresh the custodians' memory of their duty to preserve files/documents pertaining to the matter. In some embodiments, once a preservation notice has been sent to a custodian, the electronic discovery management server 110 may undertake to prevent any reimaging or refreshing of the custodian's personal computer(s) by transmitting an alert of the preservation notice to the enterprise's information technology management group. In addition, the survey responses received from custodians serve to inform the collection process. For example, one survey may inquire as to what network storage devices the custodian uses when storing data. The answer that the custodian gives to the survey may inform the addition of Shared Drives to the custodian profile in the Unified Database 122 that may be used later in collection.

[00111] According to some embodiments of the present invention, the e-discovery manager may utilize the user interface 118 to add attachments, notes, tasks, and search terms to a case or matter. In some embodiments, the contacts/managers for a case may also access the case

profile in the case database 124 using a web browser and may add attachments, notes, tasks, and search terms to be stored therein. Thus, the e-discovery manager may not be the only entry with access to the case and case management applications of the electronic discovery management server 110. The subject matter of the attachments, notes and tasks could be anything relevant to the case or matter. In some embodiments, the tasks are tasks that particular custodians must complete and the electronic discovery management server 110 is configured to transmit a notice to the custodians that that the task needs to be completed, perhaps using standard email functions. With regard to attachments, the e-discovery manager, or the contact/manager of the case, may upload relevant files to be attached to the case profile.

[00112] With regard to the search terms, the e-discovery manager or the case contacts or managers may add certain terms to the case profile to be applied when searching the collected data to locate data responsive or relevant to the underlying issues in the case. Storing the search terms within the case profile is advantageous as it creates a record of the searching that is to be undertaken with respect to the data and aids in organization of the data, as discussed further below.

[00113] According to some embodiments of the present invention, when a decision is made that it is time to collect from certain custodians in a matter, the e-discovery manager may use the user interface 118 to release the custodians from the matter to the underlying case. This release triggers the commencement of collection of the custodians' data. In some embodiments, the electronic discovery management server 110 is configured to allow all custodians assigned to the matter to be released to the case at the same time. In addition, in instances where the e-discovery manager has previously created groups of custodians within the case, the electronic discovery management server 110 is configured to allow a group of custodians to be released from a matter to a case at the same time.

[00114] DATA

[00115] Once a custodian has been identified for collection, whether manually by the e-discovery manager or by being released from a matter to a case, the electronic discovery system 100 is configured to automatically collect the custodian's data using the location information stored in the Unified Directory 122. Therefore, the electronic discovery management server 110 accesses the custodian profile of the custodian to be collected in the Unified Directory 122 and determines, from the information stored therein, the different locations of data storage for the

particular custodian that must be collected. There are many different locations that the system 100 can address, including personal computers, email accounts, and network storage areas, including HomeSpaces and Shared Drives.

[00116] If a custodian profile (for a custodian released for collection) includes at least one personal computer(s) associated with the custodian, then the electronic discovery management server 110 may undertake to collect the files on these machines. Therefore, the electronic discovery management server 110 may retrieve the relevant machine identifying information, such as domain, name, IP address, *etc.*, and may initialize deployment of a local collection application 132 running on collections server 130 (as shown in **FIG. 7**).

[00117] The local collection application 132 is configured to be deployed from the collections server 130 or another server within the network 102 to any of the enterprise personal computers 140. Therefore, for a particular custodian, the local collection application 132 is configured to utilize the machine identifying information supplied by the electronic discovery management server 110 to be deployed to the identified custodian computer. According to one embodiment, the local collection application 132 is configured to be automatically installed on the target custodian's personal computer. The local collection application 132 is further configured to generate a snapshot of the data residing on the local storage of the personal computer 140, for example, by using a commercially available application such as the Volume Shadow Copy Service, store the snapshot in a storage area on the personal computer, and transmit copies of the files included in the snapshot to the collections server 130. By transmitting the data from the snapshot of the data stored on the hard drive of the personal computer, the local collection application 132 advantageously allows the custodian to continue to use her machine without substantial interference from the local collection application 132 and even interact with the data stored on the hard drive as the snapshot of the data is being transmitted to the collections server 130.

[00118] In addition to the functions described above, the local collection application 132 may also be configured to transmit to the database server 120 a catalog of the files included in the snapshot to be stored in the ongoing collections database. This catalog may be referenced by the collections server 130 in order to determine whether collection is complete and to resume interrupted collections at the point of interruption. Additionally, in accordance with some embodiments, the local collection application 132 is configured to compile and transmit to the

electronic discovery management server 110 a list of network resources the user is using, including, for example, network applications and file servers that the user has used or accessed. This list of resources may be stored in the database server 120 in the custodian's profile in the Unified Directory 122. With regard to transmission of the files themselves, according to one embodiment of the invention, the local collection application 132 is configured to compress, hash, and upload the files included in the snapshot to the collections server 130.

[00119] In some embodiments, the electronic discovery management server 110 may utilize a computer monitoring application 117 to determine when to attempt a collection from a custodian's machine. The computer monitoring application 117 is configured to monitor the network 102 and determine which of the enterprise personal computers 140 are online. Therefore, in the event there is a custodian whose local machine needs to be collected, the computer monitoring application 117 is configured to determine when that machine joins the network 102 (*i.e.*, when it appears to the computer monitoring application 117) and inform the electronic discovery management server 110 that it should initialize the local collection application 132 immediately.

[00120] If a custodian profile (for a custodian released for collection) includes any paths for HomeSpaces or Shared Drives, then the electronic discovery management server 110 may undertake to collect the files from these file servers by initializing the file server collection application 134 running on collection server 130 (as shown in **FIG. 7**). The file server collection application 134 is configured to access the file server located at the given path, whether the file server is a HomeSpace or a Shared Drive, copy the data residing on the file server, and compress, hash, and transmit the copied data to the collections server 130. The file server collection application 134 may be programmed with preset instructions that allow it to only copy files meeting certain criteria, for example, files that have certain file extensions. Alternatively, the programmed instructions may prevent the file server collection application 134 from copying files having certain file extensions or other attributes. Either of the foregoing is advantageous if the e-discovery manager is not interested in copying executable files or source code, for example. In some embodiments, the file server collection application 134 is also configured to generate a size estimate of the files residing on the targeted file server. In one embodiment, the file server collection application 134 may automatically begin the collection process (copying and transmitting data) if the size estimate falls below a predetermined

threshold. In addition, in some embodiments, the file server collection application 134 is configured to determine whether a particular folder that it is collecting from a file server includes more than a token amount of nearline files, and, in the event that the folder does include such nearline files, choose to not collect such files so as to avoid overloading the server. Therefore, according to different embodiments, the file server collection application 134 copies all or a portion of the files residing on a file server located at the path given in the released custodian's profile and transmits them to the collections server 130.

[00121] If a custodian profile (for a custodian released for collection) includes an email address for an email account on the enterprise email server 160, then the electronic discovery management server 110 may undertake to collect the files from the enterprise email server 160 by initializing the active email collection application 136 running on collections server 130 (as shown in **FIG. 7**). In some embodiments, the active email collection application 136 is configured to access the particular Microsoft Exchange server within the enterprise email server 160 on which the custodian's account resides (which is known based on the information included in the Unified Directory 122), copy all email located there, including emails deleted by the custodian up to a predetermined period of time prior to the collection, (for example, seven days prior to the collection) and transmit the copied emails to the collections server 130.

[00122] Regardless of the storage resource location from which data is being collected, or the particular type of data being collected, the collections server 130 is configured to store the data first (while the collection is still ongoing) in the short-term staging drive 180 until the particular collection is complete, attach a barcode to the set of data resulting from the particular collection, and then copy the data set to the long-term storage area network 190 for permanent storage. Furthermore, the collections server 130 transmits the barcode information to the electronic discovery management server 110 to be stored in the database server 120, for example, in the custodian's profile in the Unified Database 122, in relation to the stored information about the particular collection, whether it was a local collection, an active email collection, a file server collection, *etc.* Therefore, the barcode can be used for reference at a later date to determine the origin of the data. After the data has been copied to the long-term storage area network 190, the collections server 130 compares the hashing of the data in permanent storage to the original data in the staging drive 180 and, if the hashing is identical, purges the data from the staging drive 180.

[00123] Once the data has entered the long-term storage area network 190, it is not necessarily ready for review. Indeed, it is likely that the data may need to be processed before it is searchable and suitable for review by investigators and attorneys. For example, the files may be encrypted in the form in which they are collected and sent to the long-term storage area network 190. Therefore, according to some embodiments, the data may be copied to the conversion services server 170 where a series of decryption and standardization functions may be applied to it. After the data is decrypted and standardized, it is returned to the long-term storage area network 190 and may remain there to be accessed for review purposes.

[00124] With reference now to **FIG. 8**, a block diagram is provided that illustrates the electronic discovery management structure of the present invention, according to some embodiments. As illustrated in **FIG. 8**, certain processes described herein may be categorized within one of case management, as represented by Block 200, custodian management, as represented by Block 220, or data management, as represented by Block 240. As described above, the electronic discovery system 100 is arranged such that cases, custodians and data may be managed independent of one another. However, there is still an element of the categorization of processes within the categories that is conceptual, and it should be understood that certain processes may be correctly assigned to more than one category. Therefore, while the architecture of the system 100 allows separate management of custodians, cases, and data, certain processes of the present invention may affect more than one of the foregoing.

[00125] The first process that falls within the case management category is creation of a matter or case as a framework for litigation support activities, as shown in Block 202. As described above, the e-discovery manager may enter into the user interface 118 certain information about a particular matter or case, such as a case name and/or number, a short description of the matter/case, a legal identifier, the particular requester (*i.e.*, who asked for the case to be opened), managers or contacts for the matter (*i.e.*, individuals involved in the substance of the matter rather than the process, like the e-discovery manager) etc.

[00126] It is noted that custodian information is stored separately from the case information allowing for the same custodian in multiple cases. This provides for the electronic discovery system of the present invention to have scalability, whereby evidence associated with one custodian may be used in multiple cases.

[00127] The electronic discovery management server 110 stores this information in the case database 124 in the database server 120. The case database 124 houses this information such that all information relating to a particular matter or case is related within the case database 124 and a user, such as a manager or contact, can use the user interface 118 to view and edit a profile of the matter or case.

[00128] The next process within case management is the creation of preservation notices and surveys specific to the matter, as shown in Block 204. In this regard, the e-discovery manager may, through the user interface 118, either generate a new preservation notices or surveys relevant to the particular case or matter to be stored in the case profile in the case database 124 or, alternatively, link a preservation notice or survey already stored in the database server 120 to the case profile of the specific case or matter at issue. Also within case management is the creation of search terms pertinent to the case, as represented by Block 206. As described above, the e-discovery manager or a contact or manager for the case may use the user interface 118 to input individual search terms or search term sets to be applied to the data harvested in the case. In some embodiments, the search terms may be limited to be used with particular custodians and/or with particular harvested data types. The search terms will be saved in the case database 124 so that they may be readily applied to harvested data and used in connection with storing the resulting responsive data.

[00129] The processes of entering relevant attachments, notes and updates to a particular case or matter also falls within the case management category, as demonstrated by Blocks 208 and 210. The e-discovery manager or a case contact or manager may use the user interface 118 to upload documents and enter notes and other relevant data, including updates and reminders, to be stored in the case profile of the case in the case database 124. Once these attachments, notes and updates are added, they may be referenced whenever a user views the case profile through the user interface 118. The cost estimation modules of the present invention are also processes that are categorized as case management processes, as shown in Block 212. In this regard, the electronic discovery management server 110 utilizes a cost estimation application to determine the cost of harvesting and reviewing data, based on a number of factors including, for example, number of custodians, amount of harvested data, data types, etc. Finally, case management also includes a number of tasking and workflow processes that are represented by block 214.

[00130] Moving now to custodian management, certain processes falling within the category of custodian management are shown in Block 220. While the processes involving generation of the Unified Directory 122 certainly could be categorized as custodian management, the processes shown in **FIG. 8** include those processes involving management of custodians within the scope of a case or matter. In that regard, the first process of custodian management included in **FIG. 8** is the addition of custodians to a case or matter, as shown in Block 222. As described above, the e-discovery manager may use the user interface 118 to link a custodian's profile from the Unified Directory 122 to the particular case profile in the case database 124. Thus, the custodian profile and case profile are correlated. The next processes within custodian management is the transmission of preservation notices and surveys to custodians, as shown in Block 224, and the presentation of the surveys to custodians, as shown in Block 226. The electronic discovery management server 110 uses the contact information in the custodian's profile in the Unified Directory 122 to transmit the preservation notice(s) and survey(s) stored in the case profile to the custodian. In some embodiments, a standard email function is used, so that the only information needed from the Unified Directory 122 is the custodian's email address. When the custodian checks her email, the survey will appear as a message therein, and when she opens that message, the survey will be presented to her. The survey may be configured such that when she fills it out, the survey is automatically transmitted back to the database server 120 for storage in the case profile and the custodian's profile.

[00131] Also falling within custodian management is the process of releasing custodians from a matter to a case, as shown in Block 228. The e-discovery manager uses the user interface 118 to mark the custodian's profile so that the custodian is now activated for collection of data. This may occur within the case database 124 since the custodian's profile is linked thereto. Once the custodian is released/marked, the electronic discovery management server 110 may access the custodian's profile and initialize collection based on the various data storage locations identified in the profile. Therefore, as represented by Block 230, the electronic discovery management server 110 may automatically determine the data types and locations of data to be harvested by accessing the custodian's profile in the Unified Directory 122. Alternatively, the e-discovery manager may manually make the same determination by accessing and viewing the custodian's profile. Finally, as with case management, custodian management also includes a number of tasking and workflow processes that are represented by Block 232.

[00132] The last category is data management, represented by Block 240. One major set of processes within data management are the processes relating to the harvesting of data, as shown in Block 242. These processes include the collection of data from all the different storage areas of a particular custodian, including the custodian's local storage on her personal computer(s), the custodian's network storage areas, the custodian's email, and any other areas, as are described herein. All of the data in the various storage areas is copied and transmitted to the collections server 130, as described in detail for each particular collection application or process. Upon reaching the collections server 130, data resulting from a particular collection is temporarily stored in the short-term staging drive 180 until the collection is complete, at which point it is stored in the long-term storage area network 190 in association with a specific identifying barcode. The foregoing process is represented by Block 244. The data may require decryption or standardization functions to be applied to it in order for it to be searchable and/or otherwise usable, so the next process that falls within data management is the copying of the data to the conversion services server 170 for analysis and conversion as necessary, as shown in Block 246. Once the data is converted, it is returned to the long-term storage area network 190 to be used in review.

[00133] Also falling within data management is the association of particular data sets with particular sets of search terms stored in the case profile of the case database 124. In this regard, certain search terms stored in the case profile are stored with the intention of being applied to certain types of data and/or certain custodian's data. Alternatively, certain search terms may be applied to all data collected for a specific case. In either instance, the electronic discovery management server 110 accesses the case profile, determines the search terms to be applied, and associates the search terms with the barcode of the appropriate data sets in long-term storage. Thus, the search terms will be applied to that data and the results will be generated and presented to reviewers for analysis. Finally, as with the other management categories, data management also includes a number of tasking and workflow processes that are represented by Block 250.

[00134] With reference to **FIG. 9**, an exemplary process for managing a case is provided, in accordance with one embodiment of the present invention. As represented by Block 302, a case or matter is created by the e-discovery manager and stored in the case database 124. Next, custodians are added to the case, as shown in Block 304, by linking the custodian profiles of the Unified Directory 122 to the case profile. Next, as represented by Block 306, the e-discovery

manager and/or the case contact or manager adds search terms to be applied to data harvested for the case, including instructions as to applying the search terms to particular data types or custodians. Block 310 represents the determination that must be made as to whether there is a matter or just a case. If there is no matter because preservation notices are not required, for example, for an audit, then the process will move straight to the initialization of data collection. On the other hand, if there is matter, rather than just a case, then the creation of preservation notices is required, as shown in Block 312.

[00135] The preservation notice, as shown in Block 314 is transmitted to the custodians added to the matter, perhaps using email. As shown in Block 316, a reminder notice module may be employed. As shown in Block 318, the reminder notice module transmits periodic reminder notices to custodians. The notices may be sent over email and may remind custodians about the preservation notice and/or remind custodians to fill out surveys. With regard to surveys, in the event a survey is required or desired, according to Block 320, a survey is created. The survey may be saved in the case profile in the case database 124. As shown in Block 322, it is possible to enable the survey to be attached to and transmitted with the preservation notices.

[00136] Next, as shown in Block 324, the e-discovery manager may release custodians from the matter to the case, which initialized collection of the custodian's data. As shown in Block 326, the e-discovery manager or the electronic discovery management server 122 accesses the custodian profile, determines the data types and location to be collected, and initializes the applicable collection applications to go collect the data. Once the data has been collected and a unique barcode has been assigned to each dataset based on the particular custodian and storage location from which it originated, as shown in Block 328, the search terms previously stored in the case profile may be assigned to the dataset based on the input instructions regarding the search terms. These search terms may be applied to the dataset and the results saved to be presented to reviewers for analysis.

[00137] With reference to **FIG. 10**, an exemplary process for managing a custodian is provided, in accordance with one embodiment of the present invention. First, as represented by Block 402, a custodian is added to a matter or case. In this regard, the custodian's profile in the Unified Directory 122 is linked to the relevant case or matter profile. In order to locate the custodian's profile, a custodian search module may be employed, as shown in Block 404. Therefore, the e-discovery manager may enter any identifying information about the custodian,

whether it is the custodian's name, network user identification, email address, etc. The custodian search module will take the input information and search the Unified Directory 122 for a match. If more than one match is obtained, the user interface 118 will present all matches and allow the e-discovery manager to browse the associated profiles to determine the intended custodian. In this way, the correct custodian is identified and the profile of that custodian is linked to the appropriate case or matter.

[00138] As represented by Block 406, the electronic discovery management server 110 may determine whether the particular custodian added is a member of the enterprise "do-not-call list." In this regard, there may be an indication in the custodian's profile in the Unified Directory 122 that the particular custodian should not be contacted regarding collections, and an alternative contact should be used, such as an administrative assistant of the custodian. Alternatively, there may be a separate do-not-call list stored in the database server 120 that must be accessed and searched to determine whether or not the custodian appears on that list. In either instance, a determination is made as to whether or not the custodian should be directly contacted, and in the event the custodian should not be directly contacted, the contact information for the custodian's assistant (or other stand-in) should be obtained. This information will be used later for transmitting preservation notices and surveys.

[00139] Next, in accordance with Block 408, a determination is made by the electronic discovery management server 110 as to whether the custodian has been added to a matter or a case. If it is a case, then the custodian is verified, as shown in Block 424, supplemental data may be added to the custodian profile in the Unified Directory 122 as required, as shown in Block 426, and then the various collection applications are initialized by the electronic discovery management server 110 for collection of the custodian's data, as shown in Block 428. On the other hand, if it is a matter, then preservation notices are required. Therefore, as shown in Block 410, a preservation notice is sent via email to the custodian or custodian stand-in. As shown in Block 412, the custodian may then be inactivated from the case because, for some reason, data does not need to be collected from the custodian. In the future, when it comes time to collect from the custodian, the custodian will be reactivated, as shown in Block 422.

[00140] After a preservation notice is sent, a determination is made by the electronic discovery management server 110 as to whether a survey is required, as shown in Block 414. It should be noted that in alternate embodiments the decision on whether to send a survey may be

made prior to sending the preservation notice. In such alternate embodiments, if the survey is required, it may become a component of the preservation notice and, thus, accessed simultaneously by the custodian. If a survey is required, it is transmitted in conjunction with a preservation notice, and the answers are collected by the electronic discovery management server 110 and stored in the database server 120, as shown in Block 416. Reminder notices for the preservation notices and surveys may also be transmitted to the custodian, as shown in Block 420. Next, once it is time to collect data, the custodian is released from the matter to the case, as shown in Block 418, and the various collection applications are initialized by the electronic discovery management server 110 for collection of the custodian's data, as shown in Block 428. In this process, the custodian's profile in the Unified Directory 122 is accessed in order to determine the various locations where the custodian may have stored data. Finally, as shown in Block 430, the custodian's data is collected.

[00141] Referring to **FIGS. 11** and **12**, flow diagrams are presented of a method 500 for harvest data from various data sources, in accordance with embodiments of the present invention. At Event 502, the collection of data ensues and, at Event 504, the type of data is identified. Data Block 506 signifies active email that is collected from an exchange system or the like. At Event 508 the automated active email collection application is implemented to collect email from identified email address. As previously noted, and in accordance with present embodiments of the invention, if a custodian profile (for a custodian released for collection) includes an email address for an email account on the enterprise email server (160), then the electronic discovery management server (110) may undertake to collect the files from the enterprise email server (160) by initializing the active email collection application (136) running on collections server (130). In some embodiments, the active email collection application (136) is configured to access the particular Microsoft Exchange server within the enterprise email server 160 on which the custodian's account resides (which is known based on the information included in the Unified Directory 122), copy all email located there, including emails deleted up to a designated prior period, for example, seven days prior to the collection, and transmit the copied emails to the collections server (130). The email collection application is also capable of implementing bulk requests and for collecting email on a scheduled basis, such as daily. The email collection application is additionally capable of being implementing

enterprise-wide and requires no server identifiers or the like to collect the active email. In this regard, the email collection application (136) serves to reduce security risk.

[00142] At Event 510, a barcoding application is implemented at a staging location, such as short-term staging drive (180) to attach a barcode to the set of email resulting from the particular collection. The barcoded data is then copied and communicated to the long-term storage area network (190) for permanent storage. Furthermore, the collections server (130) transmits the barcode information to the electronic discovery management server (110) to be stored in the database server (120), for example, in the custodian's profile in the Unified Database (122), in relation to the stored information about the particular collection. Therefore, the barcode can be used for reference at a later date to determine the origin of the data. After the data has been copied to the long-term storage area network (190), the collections server (130) compares the hashing of the data in permanent storage to the original data in the staging drive (180) and, if the hashing is identical, purges the data from the staging drive (180). As such, barcoding is performed without the need to execute the barcoding application on an exchange server and, as such no human intervention is needed in the barcode process. In accordance with embodiments of the present invention, one barcode may be assigned per custodian, per data type and per event (*i.e.*, case, matter, *etc.*)

[00143] At Event 512, the collected email data may be associated with a specific search term set or sets. When the search terms are applied, a listing of the files and documents including those terms (the "search term hit list") are presented to the reviewer and also stored in the database server (120). The reviewer may provide an indication of this to the electronic discovery management server 110, which may then make a determination that other documents within the search term hit list are more likely to be responsive.

[00144] At Event 514, the collected and barcoded active email data is copied to a processing drive for subsequent analysis. It should be noted that the nature of email data obviates the need to perform conversion and/or decryption on the data set. At Event 516, the active email data set is loaded into the analysis application and, at Event 518, the data set is exported to the requestor/reviewer for analysis.

[00145] Data Block 520 signifies other non-exchange server based email, such as email accessed through a client-server, collaborative application, such as Lotus Notes® or the like. At Event 522, NSF files or any other file types associated with non-exchange server based email is

manually harvested from an enterprise-grade email server having collaborative capabilities, such as a Lotus Domino server or the like.

[00146] At Event 522, a barcoding application is implemented at a staging location, such as short-term staging drive (180) to attach a barcode to the set of non-exchange server email resulting from the particular collection. The barcoded data is then copied and communicated to the long-term storage area network (190) for permanent storage. Furthermore, the collections server (130) transmits the barcode information to the electronic discovery management server (110) to be stored in the database server (120), for example, in the custodian's profile in the Unified Database (122), in relation to the stored information about the particular collection. Therefore, the barcode can be used for reference at a later date to determine the origin of the data. After the data has been copied to the long-term storage area network (190), the collections server (130) compares the hashing of the data in permanent storage to the original data in the staging drive (180) and, if the hashing is identical, purges the data from the staging drive (180).

[00147] At Event 526, the collected non-exchange server email data may be associated with a specific search term set or sets. When the search terms are applied, a listing of the files and documents including those terms (the "search term hit list") are presented to the reviewer and also stored in the database server (120). The reviewer may provide an indication of this to the electronic discovery management server 110, which may then make a determination that other documents within the search term hit list are more likely to be responsive.

[00148] At Event 528, the NSF files or any other file types associated with non-exchange server based email that may be encrypted is decrypted using a decryption application, in accordance with embodiments of the present invention. The encryption of NSF files occurs at the user level and, therefore only the user has the password necessary for decryption. The decryption application allows for decryption of the NSF file-type data without the knowledge of the user/encrypter. The decryption application finds ID files that exist anywhere in the enterprise system, creates a database of the ID files, associates the database with the user/encrypter and subsequently decrypts the data.

[00149] At Event 530, the non-exchange server email data set is loaded into the analysis application and, at Event 532, the data set is exported to the requestor/reviewer for analysis.

[00150] Data Block 534 signifies journaled data, such as electronic commerce data stored on a repository for the purpose of regulation, compliance to regulating bodies, such as the

Securities and Exchange Commission (SEC) or the like. At Event 536, criteria is extracted from input system and manually entered in a designated third party system for data retrieval.

[00151] At Event 538, the barcoding application is implemented at a staging location, such as short-term staging drive (180) to attach a barcode to the set of journaled data resulting from the particular collection. The barcoded data is then copied and communicated to the long-term storage area network (190) for permanent storage. At Event 540, the collected and barcoded journaled data may be associated with a specific search term set or sets.

[00152] At Event 542 source-to-processing is implemented to insure that any loose files are properly formatted in a standardized format. In this regard, according to one embodiment of the invention, loose files are examined for relevancy and, if relevant, stored in a proper data format, such as a PST file or the like. The metadata associated with the non-standardized files is retained and remains with the reformatted data files. Source-to-processing file conversions may be required on EML formatted files, MSG formatted files and the like.

[00153] At Event 544, the journaled data set is loaded into the analysis application and, at Event 546, the journaled data set is exported to the requestor/reviewer for analysis.

[00154] Referring to **FIG. 12**, data block 548 signifies data from a local Personal Computer (PC), such as enterprise PC (140). At Event 550, the local collection application (132) is implemented to collect data from designated PCs by taking a "snapshot" of the device's hard drive. According to one embodiment of the invention, the local collection application may be autodeployed thus, obviating the need for any manual entry by the e-discovery manager or the like. In other embodiments of the invention, the local collection application (132) may be employed to collect data from network storage.

[00155] At Event 552, the barcoding application is implemented at a staging location, such as short-term staging drive (180) to attach a barcode to the set of local PC data resulting from the particular collection. The barcoded data is then copied and communicated to the long-term storage area network (190) for permanent storage. At Event 554, the collected and barcoded local PC data may be associated with a specific search term set or sets.

[00156] At Event 556 source-to-processing is implemented to insure that any loose files are properly formatted in a standardized format. In this regard, according to one embodiment of the invention, loose files are examined for relevancy and, if relevant, stored in a proper data format, such as a PST file or the like. The metadata associated with the non-standardized files is

retained and remains with the reformatted data files. Source-to-processing file conversions may be required on EML formatted files, MSG formatted files, IPD formatted files and the like.

[00157] At Event 558, the local PC files that may be encrypted are decrypted using a decryption application, in accordance with embodiments of the present invention. The decryption application allows for decryption of the PC files data without the knowledge of the user/encrypter. The decryption application finds ID files that exist anywhere in the enterprise system, creates a database of the ID files, associates the database with the user/encrypter and subsequently decrypts the data.

[00158] At Event 560, the local PC data set is loaded into the analysis application and, at Event 562, the local PC data set is exported to the requestor/reviewer for analysis.

[00159] Data block 564 signifies data from network storage, such as a shared drive or HomeSpace. At Event 566, the file server collection application (134) is implemented to automatically collect data from shared drives and/or HomeSpace. According to one embodiment of the invention, the file server collection application (134) may be autodeployed thus, obviating the need for any manual entry by the e-discovery manager or the like.

[00160] At Event 568, the barcoding application is implemented at a staging location, such as short-term staging drive (180) to attach a barcode to the set of network storage data resulting from the particular collection. The barcoded data is then copied and communicated to the long-term storage area network (190) for permanent storage. At Event 570, the collected and barcoded network storage data may be associated with a specific search term set or sets.

[00161] At Event 572 source-to-processing is implemented to insure that any loose files are properly formatted in a standardized format. In this regard, according to one embodiment of the invention, loose files are examined for relevancy and, if relevant, stored in a proper data format, such as a PST file or the like. The metadata associated with the non-standardized files is retained and remains with the reformatted data files. Source-to-processing file conversions may be required on EML formatted files, MSG formatted files, IPD formatted files and the like.

[00162] At Event 574, the network storage files that may be encrypted are decrypted using a decryption application, in accordance with embodiments of the present invention. The decryption application allows for decryption of the network storage data without the knowledge of the user/encrypter. The decryption application finds ID files that exist anywhere in the

enterprise system, creates a database of the ID files, associates the database with the user/encrypter and subsequently decrypts the data.

[00163] At Event 576, the network storage data set is loaded into the analysis application and, at Event 578, the network storage data set is exported to the requestor/reviewer for analysis.

[00164] Data block 580 signifies electronic data for forensics. At Event 582, a forensic collector application, such as EnCase® may be executed on the devices of interest to collect data. According to one embodiment of the invention, the forensic collector application may be automatically deployed on the device of interest without the knowledge of the device user. In accordance with another embodiment of the invention, a computer monitoring application may be implemented (not shown in **FIGS. 11** or **12**) that monitors the network to determine the addition or subtraction of computers to the network based on network status indicators, such as ID's/IP addresses returned from the network.

[00165] At Event 584, the barcoding application is implemented at a staging location, such as short-term staging drive (180) to attach a barcode to the set of forensic data resulting from the particular collection. The barcoded data is then copied and communicated to the long-term storage area network (190) for permanent storage. At Event 586, the collected and barcoded forensic data may be associated with a specific search term set or sets.

[00166] At Event 588 source-to-processing is implemented to insure that any loose files are properly formatted in a standardized format. In this regard, according to one embodiment of the invention, loose files are examined for relevancy and, if relevant, stored in a proper data format, such as a PST file or the like. The metadata associated with the non-standardized files is retained and remains with the reformatted data files. Source-to-processing may be required on EML formatted files, MSG formatted files, IPD formatted files and the like.

[00167] At Event 590, the forensic files that may be encrypted are decrypted using a decryption application, in accordance with embodiments of the present invention. The decryption application allows for decryption of the network storage data without the knowledge of the user/encrypter. The decryption application finds ID files that exist anywhere in the enterprise system, creates a database of the ID files, associates the database with the user/encrypter and subsequently decrypts the data.

[00168] At Event 592, the forensic data set is loaded into the analysis application and, at Event 594, the network storage data set is exported to the requestor/reviewer for analysis.

[00169] Data block 596 signifies collaborative data, such as data residing at discovery sites, for example LiveLink® or the like. At Event 598, a discovery site collector application, such as a LiveLink® collector application may be executed on the devices of interest to collect data. According to one embodiment of the invention, the discovery site collector preserves at least a portion of the discovery site database in the e-discovery database, including all files and all revisions of the files. In this regard, the discovery site collector application queries against the database to define what files need to be retrieved, then copies those files based on the result of the query. Metadata pertaining to the files is retained in the case management system tables. In accordance with another embodiment of the invention, the discovery site collector application collects the documents and the related metadata and uses the metadata to automatically rename the files.

[00170] At Event 600, the barcoding application is implemented at a staging location, such as short-term staging drive (180) to attach a barcode to the set of discovery site data resulting from the particular collection. The barcoded data is then copied and communicated to the long-term storage area network (190) for permanent storage. At Event 602, the collected and barcoded discovery site data may be associated with a specific search term set or sets.

[00171] At Event 604 source-to-processing is implemented to insure that any loose files are properly formatted in a standardized format. In this regard, according to one embodiment of the invention, loose files are examined for relevancy and, if relevant, stored in a proper data format, such as a PST file or the like. The metadata associated with the non-standardized files is retained and remains with the reformatted data files. Source-to-processing may be required on EML formatted files, MSG formatted files, IPD formatted files and the like.

[00172] At Event 606, the discovery site data set is loaded into the analysis application and, at Event 608, the discovery site data set is exported to the requestor/reviewer for analysis.

[00173] Thus, present embodiments herein disclosed provide for improvements in electronic discovery. Embodiments herein disclosed provide for an enterprise-wide e-discovery system that provides for data to be identified, located, retrieved, preserved, searched, reviewed and produced in an efficient and cost-effective manner across the entire enterprise system. In addition, by structuring management of e-discovery based on case/matter, custodian and data and providing for linkage between the same, further efficiencies are realized in terms of identifying,

locating and retrieving data and leveraging results of previous e-discoveries with current requests.

[00174] Specific embodiments previously disclosed provide for determining suggestive potential custodians to be added to a case/matter in an electronic discovery system. The determination is based on a comparison of data, such as personnel data or like, associated with individuals already added to the case and all other individuals in the enterprise, who form the group of potential custodians. In addition, the comparison of data occurs over the specified timeframe defined by the case to insure that the suggested potential custodians are within the confines to the timeframe of the case. Additional embodiments of the invention provide for determining preservation notice recipients based on a comparison of data, such as personnel data or the like, associated with individuals that have already been identified to receive the preservation notice and other custodians and/or potential custodians. In addition, the determination takes into account the preservation notice history of the custodian and/or potential custodian being considered as a recipient of a specific preservation notice

[00175] While certain exemplary embodiments have been described and shown in the accompanying drawings, it is to be understood that such embodiments are merely illustrative of and not restrictive on the broad invention, and that this invention not be limited to the specific constructions and arrangements shown and described, since various other changes, combinations, omissions, modifications and substitutions, in addition to those set forth in the above paragraphs, are possible.

[00176] Those skilled in the art may appreciate that various adaptations and modifications of the just described embodiments can be configured without departing from the scope and spirit of the invention. Therefore, it is to be understood that, within the scope of the appended claims, the invention may be practiced other than as specifically described herein.

CLAIMS

1. A method for suggesting potential custodians for a case in an enterprise-wide electronic discovery system, the method comprising:

receiving, at computing device, first inputs that define a case including a timeframe associated with the case;

receiving, at a computing device, second inputs that add one or more potential custodians to the case;

determining, via computing device processor, a plurality of suggested potential custodians, wherein the determination is based on (1) a predetermined threshold of matches between predetermined data associated with at least one previously added potential custodian and a suggested potential custodian and (2) the timeframe associated with the case; and

providing, at a computing device, an output display of the determined plurality of suggested potential custodians that meet the predefined threshold of matches;

receiving, at a computing device, third inputs that select, from the output display, one or more of the plurality of suggested potential custodians and add the selected suggested potential custodians to the case; and

refining, at a computing device, in real-time, the output display of the determined plurality of suggested potential custodians based on receipt of the third inputs that select one or more the plurality of suggested potential custodians, wherein refining includes decreasing or increasing a number of the suggested potential custodians provided in the output display.

2. The method of claim 1, wherein determining further comprises accessing, via a computing device processor, one or more enterprise databases to determine the at least one match between the predetermined data associated with at least one added potential custodian and a suggested potential custodian.

3. The method of claim 1, wherein the predetermined data includes employment chain-of-command.

4. The method of claim 1, wherein the predetermined data includes employment locale.

5. The method of claim 1, wherein the predetermined data includes employment classification.

6. The method of claim 1, wherein the predetermined data includes data source accessibility.

7. The method of claim 1, wherein the predetermined data includes data content generated by potential custodians.

8. The method of claim 1, wherein determining, further comprises applying a predetermined relevancy weight value to each match of the predetermined data and summing the predetermined relevancy weight values to result in a suggested potential custodian score.

9. The method of claim 8, wherein providing an output of the determined plurality of suggested potential custodians further comprises providing a ranked listing of the suggested potential custodians, wherein the ranked listing is ranked based on the suggested potential custodian score.

10. The method of claim 1, further comprising determining, at a computing device, one or more suggested recipients of a preservation notice from amongst potential custodians previously added to the case, wherein the determination is based on an association with the one or more previously identified recipients of the preservation notice and previous preservation notice history of the one or more suggested recipients.

11. The method of claim 10, wherein determining one or more suggested recipients further comprises determining, via the computing device processor, at least one match between predetermined data associated with at least one of the previous identified recipients and a suggested recipient.

12. The method of claim 10, wherein determining one or more suggested recipients further comprises determining, via the computing device processor, that a predefined threshold of matches exist between the predetermined data associated with the at least one of the previous identified recipients and the suggested recipient.

13. The method of claim 10, wherein the predetermined data associated with the at least one of the previous identified recipients and the suggested recipient includes one or more of employment chain-of-command, employment locale, employment classification, data source accessibility or generated data content.

14. The method of claim 10, wherein determining one or more suggested recipients, further comprises applying a predetermined relevancy weight value to each match of the predetermined data and summing the predetermined relevancy weight values to result in a suggested potential preservation notice recipient score.

15. An apparatus for suggesting potential custodians for a case in an enterprise-wide electronic discovery system, the apparatus comprising:

a computing platform including a memory and at least one processor;

an electronic discovery manager application stored in the memory, executable by the at one processor and configured to receive first inputs that define a case and include a timeframe associated with the case and receive second inputs that add one or more potential custodians to the case; and

a suggested potential custodian application stored in the memory, executable by the at least one processor and including a suggested potential custodian routine configured to:

determine a plurality of suggested potential custodians for the case, wherein the determination is based on (1) a predefined threshold of matches existing between predetermined data associated with at least one previously added potential custodian and a suggested potential custodian and (2) the timeframe associated with the case, and

refine, in real time, the output display of the determined plurality of suggested potential custodians based on receipt by the electronic discovery manager application of third inputs that

select one or more the plurality of suggested potential custodians, wherein refining includes decreasing or increasing a number of the suggested potential custodians provided in the output display.

16. The apparatus of claim 15, further comprising:

a case database located in the memory and configured to store case-related data including the timeframe associated with the case; and

a custodian database located in the memory and configured to store potential custodian and actual custodian-related data,

wherein the case database is linked to the custodian database such that adding one or more potential custodians to the case is reflected in respective case-related data in the custodian database and the potential custodian-related data in the custodian database.

17. The apparatus of claim 15, wherein the suggested potential custodian routine is further configured to access one or more enterprise databases to determine the at least one match between the predetermined data associated with at least one added potential custodian and a suggested potential custodian.

18. The apparatus of claim 15, wherein the predetermined data includes one or more of employment chain-of-command, employment locale, employment classification, data source accessibility or data content generated by potential custodians.

19. The apparatus of claim 15, wherein the suggested potential custodian routine is further configured to apply a predetermined relevancy weight value to each match of the predetermined data and sum the predetermined relevancy weight values to result in a suggested potential custodian score.

20. The apparatus of claim 15, wherein the suggested potential custodian application is further configured to provide the output of the determined plurality of suggested potential custodians, wherein the output comprises a ranked listing of the suggested potential custodians, and wherein the ranked listing is ranked based on the suggested potential custodian score.

21. The apparatus of claim 15, further comprising a suggested preservation notice recipient application stored in the memory, executable by the processor and including a suggested preservation notice recipient routine configured to determine one or more suggested recipients of a preservation notice from amongst potential custodians previously added to the case, wherein the determination is based on an association with the one or more previously identified recipients of the preservation notice and previous preservation notice history of the one or more suggested recipients.

22. The apparatus of claim 21, wherein the suggested preservation notice recipient routine is further configured to determine at least one match between predetermined data associated with at least one of the previous identified recipients and a suggested recipient.

23. The apparatus of claim 21, wherein the suggested preservation notice recipient routine is further configured to determine that a predefined threshold of matches exist between the predetermined data associated with the at least one of the previous identified recipients and the suggested recipient.

24. The apparatus of claim 21, wherein the predetermined data associated with the at least one of the previous identified recipients and the suggested recipient includes one or more of employment chain-of-command, employment locale, employment classification, data source accessibility or generated data content.

25. The apparatus of claim 21, wherein the suggested preservation notice recipient routine is further configured to apply a predetermined relevancy weight value to each match of the predetermined data and sum the predetermined relevancy weight values to result in a suggested potential preservation notice recipient score.