

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SAP AMERICA, INC.,
Petitioner,

v.

CELONIS SE,
Patent Owner.

IPR2027-00004
Patent: US 11,256,708
Issued: February 22, 2022
Application No. 16/256,380
Filed: January 24, 2019

Title: METHOD OF CREATING PROCESS PROTOCOLS

DECLARATION OF CRAIG A. KNOBLOCK

SAP
EXHIBIT 1003

TABLE OF CONTENTS

	Page
I. INTRODUCTION AND ENGAGEMENT	5
II. BACKGROUND AND QUALIFICATIONS	5
III. MATERIALS CONSIDERED AND INFORMATION RELIED UPON REGARDING THE '708 PATENT	9
IV. UNDERSTANDING OF PATENT LAW	10
V. SUMMARY OF OPINIONS.....	14
VI. LEVEL OF ORDINARY SKILL IN THE ART	14
A. Person Of Ordinary Skill In The Art (“POSITA”)	14
B. State Of The Art	15
1. Database Schemas Were Regularly Used To Define Relationships Between Tables	15
2. It Was Common For Different Systems To Execute Multiple Related Processes That Generate Process Data	16
VII. THE '708 PATENT	17
VIII. CLAIM CONSTRUCTION	20
IX. GROUND 1: CLAIMS 1-14 ARE OBVIOUS OVER RINKE937 IN VIEW OF NOOIJEN AND HARRISON	21
A. Rinke937 (EX1004)	21
B. Nooijen (EX1005)	23
C. Harrison (EX1006)	26
D. Combining Rinke937, Nooijen, And Harrison	28
E. Claim 1	39
1. Element [1.1]	39

a)	Element [1.1.a].....	39
b)	Element [1.1.b]	46
c)	Element [1.1.c].....	48
d)	Element [1.1.d]	51
2.	Element [1.2].....	54
3.	Element [1.3].....	56
4.	Element [1.4].....	63
5.	Element [1.5].....	66
6.	Element [1.6].....	69
7.	Element [1.7].....	72
8.	Element [1.8].....	79
F.	Claim 2	85
G.	Claim 3	87
H.	Claim 4	89
I.	Claim 5	91
J.	Claim 6	93
K.	Claim 7	96
L.	Claim 8	99
M.	Claim 9	100
N.	Claim 10	104
O.	Claim 11	105
P.	Claim 12	106
Q.	Claim 13	108

R.	Claim 14	109
----	----------------	-----

I, Craig Knoblock, do hereby declare as follows:

I. INTRODUCTION AND ENGAGEMENT

1. I have been retained as an independent expert on behalf of SAP America, Inc. (“SAP”) to provide my analyses and opinions on certain technical issues related to U.S. Patent No. 11,256,708 (hereinafter “the ’708 patent”). I understand that this declaration will be submitted with a petition for *Inter Partes* Review (“IPR”) of the ’708 patent.

2. I am being compensated at my usual and customary rate for my time spent in connection with the IPR of the ’708 patent. My compensation is not affected by the outcome of the IPR.

3. Specifically, I have been asked to provide my opinions regarding certain prior art indicated below and whether claims 1-14 (each a “Challenged Claim,” collectively the “Challenged Claims”) of the ’708 patent would have been obvious to a person having ordinary skill in the art (“POSITA”) as of January 24, 2019. It is my opinion that each Challenged Claim would have been obvious to a POSITA after reviewing the prior art discussed herein.

II. BACKGROUND AND QUALIFICATIONS

4. I am an expert in the fields of computer science and information sciences, with particular focus on data integration, source modeling, schema alignment, entity and record linkage, information extraction, databases, and data

mining. In formulating my opinions, I have relied on my education, training, research, teaching, and professional experience in these areas. My curriculum vitae, attached as Appendix A, provides additional information concerning my qualifications, employment history, publications, and professional activities. The following summarizes some of my background and experience relevant to the '708 patent.

5. I am a Distinguished Principal Scientist of the Information Sciences Institute at the University of Southern California ("USC") and a Research Professor in USC's Computer Science Department. I joined USC and the Information Sciences Institute in 1991 and have held a succession of research and leadership positions, including Research Director of the Center on Knowledge Graphs from 2010 to 2018 and Executive Director of the Information Sciences Institute from 2018 to 2026. I received a B.S. in Computer Science with Honors from Syracuse University in 1984, an M.S. in Computer Science from Carnegie Mellon University in 1988, and a Ph.D. in Computer Science from Carnegie Mellon University in 1991.

6. My research focuses on techniques for describing, acquiring, and exploiting the semantics of data, which relates to the meaning of data and the entities and relationships it represents. My work includes modeling data sources, aligning schemas and ontologies, linking entities and records, cleaning and normalizing data, extracting information from heterogeneous sources, and combining these techniques

to create linked data and knowledge graphs. These areas are relevant to the '708 patent's subject matter, including identifying what records in source tables represent, determining relationships among tables, and using those relationships to associate information concerning an object with that object.

7. I have co-authored numerous publications concerning the modeling, transformation, and integration of structured data. Examples include “Building Data Integration Queries by Demonstration” (2007), “Semi-Automatically Mapping Structured Sources into the Semantic Web” (2012), “A Graph-Based Approach to Learn Semantic Descriptions of Data Sources” (2013), “A Scalable Approach to Learn Semantic Models of Structured Sources” (2014), and “Leveraging Linked Data to Infer Semantic Relations within Structured Sources” (2015). This work reflects my longstanding experience with understanding the structure and meaning of source data and developing methods to map and combine that data for further analysis.

8. My research also includes record linkage, entity identification, and the integration of data preparation with data mining. My publications in these areas include “Learning Blocking Schemes for Record Linkage” (2006), “Mining the Heterogeneous Transformations Between Data Sources to Aid Record Linkage” (2009), “Integration and Automation of Data Preparation and Data Mining” (2014), and “Unsupervised entity resolution on multi-type graphs” (2016). This experience

is relevant to evaluating how records identify particular objects, how information in different sources can be associated with those objects, and how source data can be prepared for analysis.

9. I also have experience generating event logs and working with and integrating real-time data. My research includes the integration of static and dynamic sources, as reflected in my coauthored publication “On-the-Fly Integration of Static and Dynamic Sources” (2013). This experience is relevant to the ’708 patent’s generation of a process protocol from source-table information identifying objects, activities, and their associated times.

10. I have taught information integration at USC repeatedly since 2002, including courses titled “Information Integration on the Web” and “Advanced Information Integration.” I have also taught courses on building knowledge graphs and geospatial data integration. In addition to my academic work, I co-founded and served as Chief Scientist of Fetch Technologies from 1999 to 2010, and Geosemble Technologies from 2004 to 2012. I have been elected a Fellow of the Association for Computing Machinery, the Institute of Electrical and Electronics Engineers, and the Association for the Advancement of Artificial Intelligence.

III. MATERIALS CONSIDERED AND INFORMATION RELIED UPON REGARDING THE '708 PATENT

11. In preparing this declaration, I have reviewed the following materials bearing Exhibit Nos. that I understand are being referenced in the above-noted IPR:

LIST OF EXHIBITS

No.	Description
1001	U.S. Patent No. 11,256,708 (“ ’708 patent ”)
1002	File History of U.S. Patent No. 11,256,708
1004	U.S. Patent App. Pub. No. 2017/0212937 to Rinke et al. (“ Rinke937 ”)
1005	Erik H.J. Nooijen et al., <i>Automatic Discovery of Data-Centric and Artifact-Centric Processes</i> , BPM 2012 Int’l Workshops, 316 (Marcello La Rosa and Pnina Soffer eds., 2013) (“ Nooijen ”)
1006	U.S. Patent App. Pub. No. 2018/0096000 to Harrison et al. (“ Harrison ”)
1007	Martti Kantola et al., <i>Discovering Functional and Inclusion Dependencies in Relational Databases</i> , International Journal of Intelligent Systems, Vol. 7 (1992) (“ Kantola ”)
1008	Yue Zhuge et al., <i>View Maintenance in a Warehousing Environment</i> , Proceedings of the ACM SIGMOD Conference, San Jose, California (1995) (“ Zhuge ”)
1009	Meihui Zhang et al., <i>On Multi-Column Foreign Key Discovery</i> , Proceedings of the VLDB Endowment, Vol. 3, No. 1 (2010) (“ Zhang ”)
1010	Will van der Aalst et al., <i>Process Mining Manifesto</i> , BPM 2011 Workshops proceedings, Lecture Notes in Business Information Processing, Springer-Verlag (2011) (“ Process Mining Manifesto ”)
1011	Claims Listing for U.S. Patent No. 11,256,708
1012	Ramez Elmasri et al., <i>Fundamentals of Database Systems</i> (7 th Ed.) (2016) (“ Elmasri ”)
1013	Oracle® Database, <i>Database Concepts</i> (Vol. 18c) (August 2018) (“ Oracle ”)

No.	Description
1014	D.A.M. Piessens, <i>Event Log Extraction from SAP ECC 6.0</i> (2011) (Master's thesis), Eindhoven University of Technology ("Piessens")
1015	Xixi Lu et al., <i>Discovering Interacting Artifacts from ERP Systems</i> , IEEE Transactions on Services Computing, Vol. 8, No. 6 (Nov./Dec. 2015) ("Lu")
1016	U.S. Patent App. Pub. No. 2017/0068705 ("Rinke705")
1017	Jan Claes et al., " <i>Merging Event Logs for Process Mining: a Rule Based Merging Method and Rule Suggestion Algorithm</i> ," Expert Systems with Applications: An International Journal, Vol. 41 (2014) ("Claes")
1018	Wil M.P. van der Aalst, " <i>Intra- and Inter-Organizational Process Mining: Discovering Processes Within and Between Organizations</i> ," 4th IFIP WG 8.1 Working Conference, PoEM 2011 (2011) ("van der Aalst")
1019	U.S. Patent App. Pub. No. 2017/0213167 to Rinke et al. ("Rinke167")

IV. UNDERSTANDING OF PATENT LAW

12. I am not an attorney. For the purposes of this declaration, I have been informed about certain aspects of the law that are relevant to my opinions. My understanding of the law was provided to me by the Petitioner's attorneys.

13. I understand that when considering the scope of the claims of a patent that the patent claim terms should generally be given the ordinary meaning that the terms would have to a person of ordinary skill in the art in question after reading the patent as of the earliest claimed priority date.

14. I understand that the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the term

appears, but in the context of the entire patent, including the other claims and the specification of the application as filed. I further understand that the principal considerations regarding the scope and meaning of the claims are the plain language of the claim (including the surrounding claim language and context), the patent specification, and the prosecution history. I understand that while a claim is to be read in light of the specification, one must generally avoid importing limitations into the claim from the specification. I am also informed that the prosecution history can often inform the meaning of the claim by demonstrating how the inventor understood the invention and whether the inventor limited the invention in the course of prosecution, making the claim scope narrower than it would otherwise be. I applied these understandings when considering the scope and meaning of the claims of the '708 patent.

15. I understand that a prior art reference anticipates an asserted claim, and thus renders the claim unpatentable, if all elements of the claim are disclosed in that prior art reference, either explicitly or inherently (i.e., necessarily present or implied).

16. I further understand that a claim is unpatentable if it would have been obvious. Obviousness of a claim requires that the claim would have been obvious from the perspective of a POSITA at the time the alleged invention was made. I understand that a claim could have been obvious from a single prior art reference or

from a combination of two or more prior art references. I understand that an obviousness analysis requires an understanding of the scope and content of the prior art, any differences between the alleged invention and the prior art, and the level of ordinary skill in evaluating the pertinent art.

17. I further understand that a claim would have been obvious if it unites old elements with no change to their respective functions, or merely substitutes one element for another known in the field, and that combination yields predictable results. While it may be helpful to identify a reason for this combination, I understand that there is no strict requirement of finding an express teaching, suggestion, or motivation to combine within the references. When a product is available, design incentives and other market forces can prompt variations of it, either in the same field or different one. If a POSITA can implement a predictable variation, obviousness likely bars its patentability. For the same reason, if a technique has been used to improve one device and a POSITA would recognize that it would improve similar devices in the same way, using the technique would have been obvious. I understand that a claim would have been obvious if common sense directs one of ordinary skill, without hindsight bias caused by knowledge of the challenged claims, to combine multiple prior art references or add missing features to reproduce the alleged invention recited in the claims.

18. I further understand that certain factors may support or rebut the obviousness of a claim. I understand that such secondary considerations include, among other things, commercial success of the patented invention, skepticism of those having ordinary skill in the art at the time of invention, unexpected results of the invention, any long-felt but unsolved need in the art that was satisfied by the alleged invention, the failure of others to make the alleged invention, praise of the alleged invention by those having ordinary skill or expertise in the art, and copying of the alleged invention by others in the field. I understand that there must be a nexus—a connection—between any such secondary considerations and the alleged invention. I also understand that contemporaneous and independent invention by others is a secondary consideration tending to show obviousness.

19. I am not aware of any allegations by the named inventors of the '708 patent or any assignee of the '708 patent that any secondary considerations tend to rebut the obviousness of any Challenged Claim of the '708 patent.

20. I understand that in considering obviousness, it is important not to determine obviousness using the benefit of hindsight derived from the descriptions found in the patent being considered.

21. I understand that Petitioner has the burden of proving unpatentability by a preponderance of evidence, which means that the claims are more likely than not unpatentable.

22. The analysis in this declaration is in accordance with the above-stated legal principles.

V. SUMMARY OF OPINIONS

23. In forming the opinions set forth herein, I have considered:

- a) The documents listed above;
- b) My education, knowledge, skills, and experience in the field of computer science and engineering; and
- c) The level of skill of a person having ordinary skill in the art at the time of the alleged invention of the '708 patent, namely January 24, 2019.

24. As I explain in further detail below, it is my professional and expert opinion that claims 1-14 of the '708 patent are obvious in view of the prior art.

VI. LEVEL OF ORDINARY SKILL IN THE ART

25. The following sections explain pertinent aspects of the state of the art and the level of ordinary skill in the art as of January 24, 2019, the effective filing date of the challenged claims.

A. Person Of Ordinary Skill In The Art (“POSITA”)

26. A person of ordinary skill in the art at the relevant time would have had a bachelor's degree in computer science, information systems, data science, or similar discipline, and approximately two years of academic, research, or industry experience with process (workflow)-mining, data-mining, or similar systems for

analyzing event, log, or process data. Additional education could substitute for professional experience and vice versa.

27. As of January 2019, I possessed at least the skill and knowledge of a POSITA. Unless otherwise stated, my opinions herein reflect the perspective of a POSITA.

B. State Of The Art

**1. Database Schemas Were Regularly
Used To Define Relationships Between Tables**

28. Relational database schemas were conventionally used to define both the structure of database tables and the relationships among those tables. As explained in Elmasri, relationships in a relational schema may be represented by corresponding primary and foreign keys in different tables, with matching key values identifying related records. EX1012, 163-64. Oracle similarly explains that keys “describe the relationships between the tables and columns of a relational database,” and Piessens explains that foreign keys link tables and that associated cardinalities indicate the nature of those relationships. EX1013, 5-3; EX1014, 10. Thus, it was well known to use primary and foreign key information, which were known aspects of relational database schema, to identify how the database tables are related to one another.

2. It Was Common For Different Systems To Execute Multiple Related Processes That Generate Process Data

29. It was well known that a single end-to-end process could span multiple organizations and information sources. Van der Aalst observed in 2011 that, although most process mining projects had focused on individual organizations, “many processes are distributed over multiple organizations and different organizations are executing similar processes,” leading the paper to specifically address process mining in such cross-organizational settings. EX1018, 1-2. Rinke705 explained that when “real processes such as business processes are carried out with the help of one or multiple IT systems, the process instances leave data trails in the participating IT systems.” EX1016, [0015]-[0017]. Rinke705 recognized that, when a process is distributed among different IT systems, the data must be loaded “from one or more source systems” for process mining. *Id.*, [0103], Fig. 1. The Process Mining Manifesto (hereinafter, “the Manifesto”), similarly recognized that event data for a process may be “distributed over a variety of sources” and described a cross-organizational process as a “jigsaw puzzle,” in which “the overall process is cut into parts and distributed over organizations” that cooperate to complete a process instance. EX1010, 14-16. The Manifesto recognized that, in such scenario, the resulting event logs from these processes distributed over different databases must be merged to discover the end-to-end process. *Id.* 15.

30. Consistent with this understanding, Claes explained that, for an inter-organizational process, historical process data are “distributed over different sources,” each containing partial information about the overall business process, and may be combined at the “raw data level” by merging databases or files. EX1017, 2. Thus, it was well known that raw process and other data for related business processes was often generated and stored across multiple different source systems that executed the processes, and that the data would be combined for analysis of an associated overarching (e.g., “cross-organizational”) business process. Consistent with this conventional understanding, the ’708 patent itself describes in its background that, “in practice,” different “partial processes” of an overall process are executed on different computer systems and their data is stored separately. EX1001, 1:43-49.

VII. THE ’708 PATENT

31. The ’708 patent issued February 22, 2022, from an application filed January 24, 2019. It describes processes executed in or aided by a computer system, such as “business processes (e.g., an order process) or technical processes (e.g., a coating process).” EX1001, 1:13-16. These processes create data that “is stored in the computer system.” *Id.*, 1:17-20. The patent uses the term “process protocol” to refer to the “processed form” of information extracted from data created by computer

processes such as “business processes” or “technical processes.” *Id.*, 1:32-37; 1:13-20.

32. The ’708 patent explains that it was “known to analyze” such processes by extracting relevant information from the data to “be analyzed in order to identify for example, bottlenecks or other anomalies in the processes.” *Id.*, 1:21-30.

33. The ’708 patent criticizes the prior art as purportedly being unable to analyze processes that were executed on multiple computer systems. *Id.*, 1:61-66. As an example, it describes a technical process where “information relating to the assembly process may be stored in a first computer system, and information relating to a subsequent technical test process may be stored in a second computer system.” *Id.*, 1:57-60. The ’708 patent admits that these sub-processes can separately “be analyzed by means of methods known from prior art,” but alleges that an analysis of the overall process “is not possible.” *Id.*, 1:61-66.

34. The patent alleges a problem with prior approaches: where process data is created and analyzed within a respective computer system, only the processes executed in that individual system can be analyzed. In practice, however, partial processes of a higher-level process may be carried out on different computer systems and may create data on those different systems. EX1001, 1:38-46.

35. The patent later provides an example of a “process protocol”:

Based on the tables listed above, for the activity Change Order, the following process protocol is obtained:

Identifier of the process	Identifier of the process step	Sequence
1	Change Price	01.01.2018
1	Change Amount	02.01.2018
1	Change Amount	03.01.2018
1	Change Price	04.01.2018
1	Change Price	05.01.2018

EX1001, 15:39-49.

36. The '708 patent provides an illustrative example of a process in which “order information” and “order history” are stored in tables in a first computer system, while vendor call data is stored in a second computer system. EX1001, 11:29-67. To address this distributed data scenario, claim 1 recites a computer-implemented method performed in a local computer system. The method imports data tables from first and second external computer systems into the local computer system; determines and reconstructs relationships between the imported tables; selects a first data table whose records represent objects for which the process protocol is created; selects second data tables whose records represent process steps; determines activities for the selected objects; creates an executable transformation based on the selected tables and relationships; and executes the transformation to extract data from the external-system tables, store the extracted data as a process protocol, and provide the protocol to a process mining system. EX1001, claim 1.

37. The specification describes a corresponding example in which data tables T1-T12 are first imported into an internal computer system. Relationships between imported tables are then reconstructed, including relationships among tables from the same external computer system and relationships between tables from different external computer systems. After reconstruction, a database scheme exists in the internal computer system that describes the imported tables and their relationships, and the system determines suitable business objects, such as orders, for creation of a process protocol. EX1001, 6:1-33.

VIII. CLAIM CONSTRUCTION

38. For this declaration, I apply the plain and ordinary meaning of all claim terms, in addition to the proposed construction as noted below. I reserve the right to offer my opinions on any constructions or construction issues presented by Patent Owner.

39. I have been asked to apply the following proposed construction:

“determining and reconstructing the relationships between the imported data tables” (claim 1): *“after importing the data tables, determining and reconstructing relationships between the imported data tables, where the relationships between tables imported from different computer systems were unknown prior to the determining and reconstructing.”* EX1002, 149, 152, 196-197.

**IX. GROUND 1: CLAIMS 1-14 ARE OBVIOUS OVER
RINKE937 IN VIEW OF NOOIJEN AND HARRISON**

40. As explained below, claims 1-14 are obvious over Rinke937 in view of Nooijen and Harrison.

A. Rinke937 (EX1004)

41. U.S. Patent Application Publication No. 2017/0212937 to Rinke et al. (“Rinke937”) published July 27, 2017, and is prior art under 35 U.S.C. § 102(a)(1). EX1004, cover.

42. Rinke937 describes methods for “sens[ing] process steps and build[ing] up event logs based on raw data stored in a source system.” *Id.*, [0010]. These methods “transform the source data into an event log format for further process mining analysis.” *Id.*; *id.*, [0002]-[0009]; [0072] (“[R]aw data from the IT source system is transformed into a representation which is capable for process mining applications.”). These event logs allow process mining tools to reconstruct and analyze the actual process executed in the source system. *Id.*, [0088].

43. Rinke937’s event logs include process element identifiers and process steps associated with each element, and reflect the order of steps, e.g., by timestamp. *Id.*, [0012]-[0023], [0075]-[0086]. Rinke937 provides example event logs, such as the following:

Unique ID	Process Step	Timestamp (Order)
1	Purchase Requisition	2012-01-01 13:05
1	Purchase Requisition Approval	2012-01-01 15:09
1	Purchase Order Item	2012-01-02 17:03
1	Goods Receipt	2012-01-03 13:01
1	Invoice Receipt	2012-01-03 15:00
2	Purchase Requisition	2012-01-05 11:00
2	Purchase Requisition Approval	2012-01-05 17:00
...		

Id., [0086].

44. In Rinke937, the transformation from raw data to event logs involves units called “process sensors,” which “derive process data from the raw data [] of the executed processes” and generate event logs “from the derived process data.” EX1004, [0072], [0080]; *id.*, [0091] (“[A] process sensor is a unit which can be applied to the raw data of the source system to sense one or multiple process steps.”). When executed, process sensor units apply transformations that extract information relating to the process that generated the raw data, such as element identifier, step identifier, and timestamp. *Id.*, [0090]-[0102]. The process sensor units may be “scripted in a programming language” or defined “through a graphical user interface” (*id.*, [0115]-[0117]). From these inputs, “executable code for the underlying execution platform is generated.” *Id.*, [0119].

45. Rinke937's process sensor units sense and extract data that is "often scattered among different tables." *Id.* [0096]; *id.*, [0099] (describing "consecutive join[s] over (possibly) multiple tables" to derive the event log information). For example, in Rinke937's "change of content" example, a unit derives an invoice identifier from table BSEG, process-step information from table CDPOS, and a timestamp from table CDHDR. *Id.*, [0130]-[0135]. The unit relates those tables and transforms their contents into a common event log. *Id.*

B. Nooijen (EX1005)

46. Nooijen, titled "Automatic Discovery of Data-Centric and Artifact-Centric Processes," was published in 2013 by Springer-Verlag. EX1005, 316. I understand that Nooijen qualifies as prior art under Section 102(a)(1). Based on its content, publisher, and publication, it is my opinion that POSITAs would have known of and had access to Nooijen by its publication date in 2013. *Id.*, (i)-(v) (2013 Copyright and catalog information); EX1015, 872 (citing Nooijen).

47. Like Rinke937, Nooijen relates to transformation of raw process data (stored in relational data tables) into an event log that can be used for process mining. EX1005, 316-319. This technique is reflected in Nooijen Fig. 1:

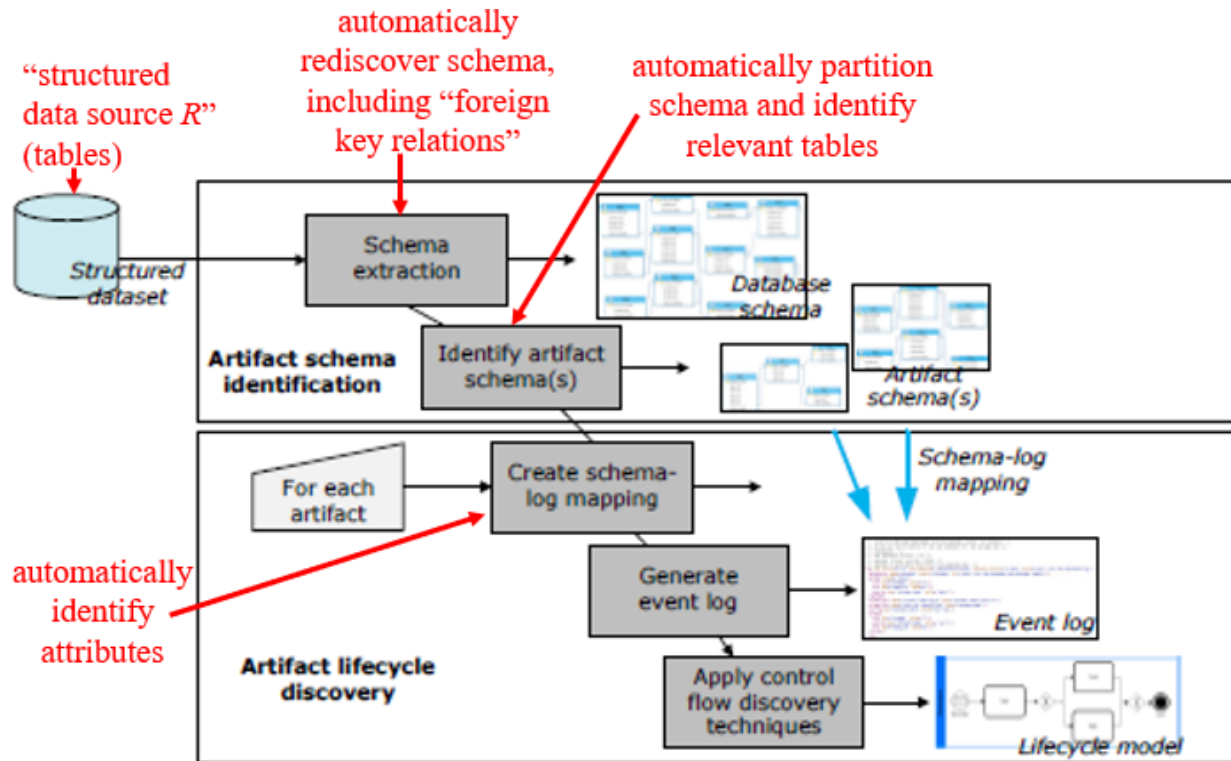


Fig. 1. Approach to Artifact Discovery

EX1005, 317 (annotated in red).

48. Nooijen seeks to improve existing approaches through techniques for “automatically discovering a process model” (*id.*, 316-317), including “automatic discovery of schema-log mapping needed for log extraction” (*id.*, 326). Nooijen explains that this avoids a manually-defined “mapping between source data and event log.” EX1005, 318.

49. Nooijen’s “automated approach” automatically reconstructs undocumented relationships among source-data tables, identifies a main table and the related tables for a process object, determines the event specification for that

object, and generates and executes database queries that transform the related table data into a timestamp-ordered event log for process mining analysis. EX1005, 318-324.

50. For a given data source, the Nooijen technique determines and reconstructs a relational schema S that represents the source tables, including column types, primary keys, and foreign key relationships. *Id.*, 319. Nooijen explains that this rediscovery may be necessary because ERP database schemas can be incomplete, including because actual foreign-key relationships may not be documented. *Id.*, 317.

51. After rediscovering and reconstructing the schema, Nooijen partitions the schema into artifact schemas. *Id.* Each artifact schema describes the information model for one artifact (business entity) and includes the tables containing information relevant to that artifact. *Id.* Nooijen also identifies a main table, T^m , for each artifact schema. *Id.*, 320. The main table's primary key serves as the identifier of the artifact, while other tables provide additional attributes for that artifact. EX1005, 319-321. Nooijen then automatically extracts an event specification, referred to as a schema-log mapping, from the artifact schema, using the selected main table. That event specification is used to construct database queries that extract events and their attributes, group the events into cases, order them by timestamp, and write the result into a classical log. EX1005, 321-324.

52. Nooijen adapts “the log extraction technique of XESame,” with an algorithm that “fully automatically” creates an “event log-trace mapping” that would otherwise be specified manually. *Id.*, 324. With that mapping, “XESame automatically generates the database queries needed to extract the log.” *Id.* Those queries extract events, informed by the reconstructed foreign-key relationships, and the resulting event traces are grouped by trace identifier, ordered by timestamp, and written to an XES format event log. *Id.*

C. Harrison (EX1006)

53. U.S. Patent Application Publication No. 2018/0096000 to Harrison et al. (“Harrison”) was published on April 5, 2018, and qualifies as prior art under 35 U.S.C. § 102(a)(1). EX1006, Cover.

54. Similar to Nooijen, Harrison is directed to identifying relationships between tables. Harrison discloses techniques for consolidating information stored in separate computer systems. It explains that related information may be stored in different source systems—for example, separate business units may maintain different databases containing information about the same customers. EX1006, [0065]. Because organizations need “to bring together data from diverse data sources to enable processing and analysis of data sets” (*id.*, [0003]), Harrison teaches a system for importing entire data tables from multiple structured sources, such as relational databases, and storing copies of those tables in a central data repository

where the imported data can be processed, managed, and analyzed. *Id.*, [0053]-[0055].

55. Harrison's "Data Tap" tool imports data "from any number of data sources" into a central repository referred to as a "data lake." *Id.*, [0055]-[0056]; *see also id.*, [0059] (describing importing tables "from multiple data sources"). External data sources may include relational databases managed by Oracle, MySQL, Microsoft SQL Server, or other database-management systems. *Id.*, [0058]. Harrison states that the tool "may import a database in its entirety (i.e. including all tables)." *Id.*, [0059].

56. Harrison stores copies of the imported tables in the central repository, which may be a Hadoop Distributed File System. *Id.*, [0056]-[0063]. Each imported file generally contains "the full set of columns imported from the source table." *Id.*, [0061] ("typically all columns of the source table will be imported"), [0089]-[0090] ("the full table contents are read" from the external source database).

57. Harrison explains that this centralized architecture permits an organization to "bring together information from many disparate databases ... and analyse [sic] and process the data centrally," rather than requiring separate analysis on each external source system. *Id.*, [0064]-[0066]. These centralized copies also preserve historical information that may later be changed or deleted in the operational source systems. *Id.*, [0063].

58. After importing and saving the data tables locally, Harrison teaches to analyze them to “automatically discover relationships” between the tables, including primary- and foreign-key relationships. EX1006, [0067]; *see also id.*, [0068], [0221]-[0225]. These relationships can be used to construct join queries that, when executed, combine and extract data from the centrally stored tables. *Id.*, [0068], [0235]; *see also id.*, [0418]-[0424].

59. Harrison and the '708 patent both arise in the technical area of gathering distributed data and preparing it for analysis. The supposed problem identified by the '708 patent involves identifying relationships among data tables originating from different computer systems or processing units, so that their data can be analyzed. *E.g.*, EX1001, 2:3-58, 12:39-13:17. Harrison directly addresses that problem with teachings on importing disparate data collections created by different operations, discovering relationships among the different source data, and using the identified relationships to facilitate analysis, *e.g.*, through queries and joins. *E.g.*, EX1006, [0064]-[0068], [0410], claim 22. A POSITA seeking to solve the problem purportedly addressed by the '708 patent would therefore logically consult Harrison's relationship-discovery techniques.

D. Combining Rinke937, Nooijen, And Harrison

60. A POSITA would have been motivated, with a reasonable expectation of success, to implement an event log generation system as disclosed by Rinke937

with Nooijen's automated relationship discovery teachings and Harrison's centralized import-and-analyze teachings. Such a system would reduce manual effort and efficiently generate event logs from process data created on multiple source systems. The resulting Rinke937-Nooijen-Harrison system would include the following operations, forming a first stage that creates an executable process-sensor unit using information derived from locally imported data, and a second stage that executes the executable process-sensor unit to generate an event log from "raw data stored in the source systems" (EX1004, [0090]) for providing to a process mining system:

- Import data tables from different source computer systems and store them in local storage, as taught by Harrison (EX1006, [0055]-[0063], [0078]-[0090]), the data tables being generated by processes running on those source computer systems, as taught by Harrison (*e.g.*, *id.*, [0064] (import from disparate databases "supporting different operations of the organization"), [0065] ("importing data" from systems for gas-supply and maintenance processes));
- Analyze the imported tables to determine and reconstruct relationships between tables, including undocumented primary-key and foreign-key relationships, as taught by Harrison (*id.*, [0067]-[0068], [0221]-[0225]),

using the automated schema discovery taught by Nooijen (EX1005, 317, 319-20);

- Select, from the imported tables, a “main table” whose records identify the individual business objects for which an event log will be created, as taught by Nooijen (*id.*, 317, 319-22);
- Select related event tables based on the main table and reconstructed relationships, and determine activities associated with each identified object, as taught by Nooijen (*id.*, 320-24);
- Create an executable transformation using the selected tables, reconstructed relationships, and determined activities, by combining Nooijen’s teachings on automatic query generation and event log creation (*e.g.*, *id.*, 324) with Rinke937’s teachings on event log generation (EX1004, [0031]-[0032], [0096]-[0102], [0116]-[0119]);
- Execute the transformation to extract object identifiers and associated activity data from the tables on the different source systems, as taught by Rinke937 (*id.*, [0012]-[0020], [0090]-[0102], [0112]-[0119]) and Nooijen (EX1005, 324), using network connection information as taught by Harrison (EX1006, [0055], [0087]-[0090], [0455]-[0456]); and

- Store the extracted data as an event log, associating each object with its process steps and their order, and provide that log to a process mining system for analysis and reconstruction, as taught by Rinke937 (EX1004, [0017]-[0026], [0035], [0075]-[0088], [0112]-[0113]).

61. POSITAs would make routine implementation choices when embodying this combination and would have made this combination for at least the following reasons.

62. *First*, a POSITA would have been motivated to use the automated discovery and mapping teachings of Nooijen to reduce the manual effort of creating the executable process-sensor units taught by Rinke937. Rinke937 explains that relevant data fields are “scattered among different tables” and that their relationships “have to be defined within the process sensor(s),” but first requires manual definition of table relationships through a GUI or programming/scripting. EX1004, [0096]; *id.*, [0097]-[0102], [0115]-[0119]. Nooijen recognizes that existing approaches require a mapping “to be defined manually,” whereas Nooijen teaches creating the mapping “fully automatically.” EX1005, 324. Specifically, Nooijen expressly identifies the downsides of manually selecting tables and defining source-to-log mappings, which are “hard and time consuming,” suffer from human bias, and are often “incomplete.” EX1005, 317-319. Nooijen addresses these issues with its automated approach that reconstructs relationships, identifies tables relevant to an artifact/object, and

generates the queries needed to extract related events. EX1005, 323-24, 326. A POSITA thus would have naturally applied those teachings from Nooijen to automate the determination of tables, relationships, and other information when creating the process sensors as taught by Rinke937, to reduce manual effort and improve accuracy. *See* EX1005, 318 (explaining that manual approaches are “time-consuming” and inaccurate). POSITAs would have understood that the automated approach of Nooijen complements and improves the Rinke937 event log generation framework by minimizing the need for time-consuming manual specification of the executable units for extracting data and generating event logs, accounting for relationships omitted from the documented source schema, and basing the transformation on the source system’s actual data rather than a user’s assumptions about the data.

63. *Second*, a POSITA would have been motivated to use the techniques taught by Harrison for importing data tables locally to improve the automated analysis taught by Nooijen. Harrison identifies a “need to bring together data from diverse data sources to enable processing and analysis” and explains that data imported from “disparate databases” is often “interrelated,” and teaches importing data from multiple data sources to “analyse and process the data centrally.” EX1006, [0003], [0064]-[0068]. This addresses a recognized circumstance in process mining: relevant process information may be distributed across different source systems and

must be combined to reconstruct the overall process. A POSITA knew that this was often the case for event logs / process protocols, as taught by Nooijen and Rinke⁹³⁷. *E.g.*, EX1010, 12 (“Process Mining Manifesto,” describing the need to merge data “distributed over a variety of sources”), 15 (processes spanning across collaborating organizations). Skilled artisans recognized that relevant process data often resides in different computer systems and that combining that data enables more complete process reconstruction. *Supra* Section VI.B.2. The Process Mining Manifesto expressly explains: “Data may be distributed over a variety of sources. This information needs to be merged.” EX1010, 12. Addressing processes spanning collaborating organizations, it further explains that “[a]nalyzing the event log within one of these organizations involved is insufficient” and that “[t]o discover end-to-end processes, the event logs of different organizations need to be merged.” *Id.*, 15; *see also* EX1017, 2. Thus, a POSITA would have understood the need to handle those common situations where process data was generated on different source systems, motivating a POSITA to use Harrison’s complementary teachings on multi-source data. EX1006, [0055], [0059], [0064]-[0065]; EX1010, 12, 15. These teachings thus would have motivated a POSITA to incorporate the import functionality taught by Harrison so that Rinke⁹³⁷-Nooijen-Harrison determines relationships and configures process sensor units using related data originating in different systems.

64. Importing the tables would make their contents available for discovering relationships across the source systems. Harrison explains that related information may reside in separately maintained databases, that related columns need not have matching names, that source documentation may be “incomplete or inconsistent,” and that relationships “may be lost during the import process,” and addresses those problems by automatically discovering relationships among imported tables. EX1006, [0065]-[0068]; *id.*, [0005], [0055], [0059]-[0064]. Harrison expressly recognizes that relationships between tables that “originated from different data sources ... may not be known a priori” and teaches discovering those relationships through analysis of the imported table contents. *Id.*, [0421]; *see also id.*, [0067]-[0068], [0221]-[0225]. Nooijen teaches complementary techniques for reconstructing schema information, including “undocumented foreign key relations,” and using the resulting schema to generate mappings for event-log extraction. EX1005, 319-324. A POSITA thus would have been motivated to apply these automated discovery techniques to locally imported tables in order to discover the relationships needed for event-log extraction across the imported data collection. EX1005, 316; EX1006, [0066]-[0068].

65. The processing demands of automated discovery provide a further reason to make the table contents locally available. Automated relationship discovery requires examining many candidate combinations of columns and

examining their contents. Kantola illustrates this problem for inclusion dependencies, which identify containment relationships between column values and can help identify candidate foreign-key relationships. For two tables with n attributes each, Kantola explains that there are more than $n!$ possible nonequivalent inclusion dependencies. EX1007, 603. Kantola also proves that a particular problem of determining whether a suitable column correspondence satisfies an inclusion dependency is NP-complete, while recognizing that practical techniques can reduce the candidate space, including comparing compatible attribute types and concentrating on shorter dependencies. EX1007, 603-604. Nooijen confirms the practical processing demands of schema discovery, reporting substantial execution times for identifying primary keys, foreign keys, and the information used for table clustering. EX1005, 325. A POSITA would have understood that making the table contents locally available permits their reuse across these comparisons without repeatedly retrieving the underlying table data from the external systems.

66. By January 2019, the efficiency benefits of analyzing locally available, integrated source information were sufficiently well established that a POSITA would have understood those benefits. Zhuge's 1995 publication reflects this longstanding understanding. Zhuge describes a data warehouse as a repository containing information extracted from distributed and heterogeneous sources, translated into a common model, and integrated for queries and analysis, including

data mining. EX1008, 316. Zhuge explains that queries and analysis can be performed efficiently because the integrated information is directly available at the warehouse. *Id.* A POSITA would have recognized the same benefit when applying the discovery techniques taught by Nooijen to the locally imported tables taught by Harrison: the analysis compares and reuses the imported contents within a common repository, reducing repeated network access and repeated queries to source databases. Local storage does not eliminate the computational work of discovery, and importing the tables itself incurs transfer costs. But those initial costs can be offset when the imported contents support many subsequent comparisons during relationship discovery and mapping. This efficiency benefit would have motivated a POSITA to use locally imported tables for the first stage of Rinke937-Nooijen-Harrison, in which the system determines relationships and creates the executable process sensor units.

67. *Third*, a POSITA would have been motivated to retain the process sensor framework taught by Rinke937 in the second stage of Rinke937-Nooijen-Harrison for generating event logs from “raw data stored in the source systems.” EX1004, [0090]. Rinke937 expressly identifies “advantages” of “using process sensors to create event logs instead of using database query scripts,” including configuration “to specifics of the source data system” and execution “at fixed time intervals.” EX1004, [0056]-[0064], [0113]. Those benefits would have motivated a

POSITA to improve the creation of the executable process sensor units taught by Rinke937 through the automated discovery of table relationships in locally imported tables as taught by Nooijen and Harrison, while continuing to use the process sensor framework taught by Rinke937 for execution in the second stage of operations. EX1004, [0090]-[0102], [0117]-[0119]; EX1005, 319-324; EX1006, [0066]-[0068], [0221]-[0225]. In particular, Rinke937 teaches that its process sensor units operate on “raw data stored in the source systems” to create event logs. EX1004, [0090]. The improved executable process sensor units of Rinke937-Nooijen-Harrison would thus be executed against the corresponding original source tables to generate the event log, as taught by Rinke937. EX1004, [0090]-[0102], [0117]-[0119]. This allows the executable process sensor unit to be reused and repeatedly executed to generate event logs as additional activities are recorded in the source systems, as taught by Rinke937. EX1004, [0063], [0090], [0113].

68. *Finally*, a POSITA would have had a reasonable expectation of success combining these teachings because Rinke937, Nooijen, and Harrison describe compatible operations and information. Nooijen describes successful use of its automated technique in a system like Rinke937, which transforms source data into event logs for process mining. EX1004, [0010], [0072], [0080], [0090]-[0102]; EX1005, 316-24. For example, both Rinke937 and Nooijen teach using table relationships as part of extracting data fields for inclusion in an event log. EX1004,

[0031]-[0032], [0095], [0103]-[0119]; EX1005, 321-24. Thus, a POSITA would have expected success in implementing a system that used Nooijen's automatic relationship determination teachings to generate relationships and other information, and then use those relationships in the executable code creation and event log generation process of Rinke937. Similarly, Harrison expressly contemplates the use of its centralized import teachings in systems that determined primary-key and foreign-key relationships between disparate data sources. EX1006, [0058]-[0068], [0221]-[0235]. Nooijen uses the same kinds of tables and relationships to generate its mappings, and Rinke937 uses table relationships to define the extraction it performs. EX1005, 317, 319-24; EX1004, [0096]-[0102]. A POSITA would have thus seen Harrison's teachings as a natural and readily implemented extension of the Rinke937-Nooijen event log generation process. Further, Harrison teaches importing tables into a conventional relational database, providing a compatible environment for the relational analysis taught by Nooijen. EX1006, [0055]-[0056]. Its metadata preserves the correspondence between imported columns and their original source databases, tables, and columns, including any column renaming. *Id.*, [0111]-[0117]. A POSITA thus would have expected to resolve the locally determined references to their external counterparts without changing the key-based associations represented by the mappings. POSITAs would have thus recognized

that the combination involved the predictable application of known database, software, and data-mining techniques according to their established functions.

69. So combined, Rinke937-Nooijen-Harrison taught and rendered obvious claims 1-14, as described in detail below.

E. Claim 1

1. Element [1.1]

70. To the extent the entire preamble has patentable weight, Rinke937-Nooijen-Harrison teaches it, as explained in the following sections.

a) Element [1.1.a]

[1.1.a] A computer-implemented method of creating a process protocol in a local computer system (LC), which comprises a processor and a storage device, wherein the process protocol encodes object related activities and is created from raw data, which is stored in part in a first external computer system (EC1) and in part in a second external computer system (EC2),

71. Rinke937-Nooijen-Harrison satisfies this element. As discussed below, Rinke937-Nooijen-Harrison teaches creating, in a local computer system, a process protocol encoding object-related activities from raw data stored in part in two external computer systems.

72. Rinke937 expressly discloses “a method for generating, in a computer system having a processor and a storage means operatively coupled to the processor, an event log from raw data stored in a source system” (*a computer-implemented method for creating a process protocol in a local computer system (LC), which*

comprises a processor and a storage device, wherein the process protocol encodes object related activities and is created from raw data). EX1004, [0012]; *id.*, [0001], [0051]-[0055]. Rinke937-Nooijen-Harrison includes the computer system taught by Rinke937 and generates, in that computer system, an event log from raw data as taught by Rinke937. *Supra* Section IX.D.

73. The event log of Rinke937-Nooijen-Harrison is a *process protocol*.

The “event log” taught by Rinke937 has a “data structure comprising at least a first attribute for storing the unique identifier of the process element of the respective process step, a second attribute for storing the identifier of the respective process step, and a third attribute for storing the order of the process steps within a process element.” EX1004, [0020]-[0023]; *id.*, [0014]-[0019] (disclosing a process sensor unit deriving the unique identifier(s), identifier(s) and order of steps from the raw data), [0005]-[0006], [0025]-[0026], [0075]-[0088], [0091]-[0102], [0124]-[0136], Fig. 5. Likewise, the “data structure of the process protocol” described by the ’708 patent “comprises at least a first attribute, in which a unique identifier of the process of the respective process step is stored, a second attribute, in which an identifier of the respective process step is stored, and a third attribute, in which a sequence of the process steps within a process is stored.” EX1001, 9:8-15; *see also id.*, 15:21-49.

74. Rinke937 “shows an example of an event log” with this data structure:

Rinke937's Event Log

Unique ID	Process Step	Timestamp (Order)
1	Purchase Requisition	2012-01-01 13:05
1	Purchase Requisition Approval	2012-01-01 15:09
1	Purchase Order Item	2012-01-02 17:03
1	Goods Receipt	2012-01-03 13:01
1	Invoice Receipt	2012-01-03 15:00
2	Purchase Requisition	2012-01-05 11:00
2	Purchase Requisition Approval	2012-01-05 17:00
...		

EX1004, [0086] (red annotations added). The '708 patent provides a table showing an example “process protocol” with similar attributes as the event log:

'708 Patent's Process Protocol

Identifier of the process	Identifier of the process step	Sequence
1	Change Price	01.01.2018
1	Change Amount	02.01.2018
1	Change Amount	03.01.2018
1	Change Price	04.01.2018
1	Change Price	05.01.2018

EX1001, 15:42-48 (red annotations added). The “event log” taught by Rinke937 is therefore a *process protocol* because its data structure includes the same attributes and information regarding a process as the “process protocol” described by the '708 patent. *See also* EX1002, 196 ('708 applicants admitting that such “event logs” in Rinke167, which shares disclosures with Rinke937, “may show parallels to the data structure of a process protocol created by the method of claim 1”).

75. **The event log of Rinke937-Nooijen-Harrison *encodes object related activities*.** Specifically, the event log taught by Rinke937 records “an unique ID, a process step, and an order of the process steps,” where the unique identifier (ID) corresponds to a “process object.” EX1004, [0081]-[0084], [0099]. Each event log entry thus associates a process step with an identified “object,” and the event log therefore *encodes object related activities*. *E.g.*, EX1004, [0124]-[0129] (providing example where the objects are invoices).

76. **The event log of Rinke937-Nooijen-Harrison is *created from raw data which is stored in part in a first external computer system and in part in a second external computer system*.** Specifically, as explained in the following paragraphs, Rinke937-Nooijen-Harrison creates an event log from “raw data stored in the source systems” as taught by Rinke937, where that raw data is stored in part in a first external computer system and in part in a second external computer system as taught by Harrison. EX1004, [0090]; EX1006, [0055], [0058]-[0059], [0065], [0456].

77. As noted above, Rinke937 discloses creating the event log (*process protocol*) “based on raw data stored in a source system, i.e., to transform the source data into an event log format for further process mining analysis.” EX1004, [0010]; *see also id.*, [0008], [0035], [0079], [0088], [0114], [0121]. Its process sensor units “derive process data from the raw data comprising the digital traces of the executed

processes” and generate the event log from that data. *Id.*, [0080]; *see also id.*, [0001], [0003], [0008], [0012]-[0023], [0072]-[0079], [0090]-[0094]. The source data is stored in database tables, and the relevant fields may be “scattered among different tables.” *Id.*, [0096]; *see also id.*, [0097]-[0102], [0124]-[0136]. Rinke⁹³⁷ further discloses without elaboration raw data in plural “source systems.” EX1004, [0090], [0103].

78. Harrison discloses importing data “from any number of data sources 102-1, 102-2, 102-3 into a data repository 108,” including “tables and data” originally stored in “multiple data sources.” EX1006, [0055], [0059]. Harrison depicts these data sources as separate and distinct in Figure 1:

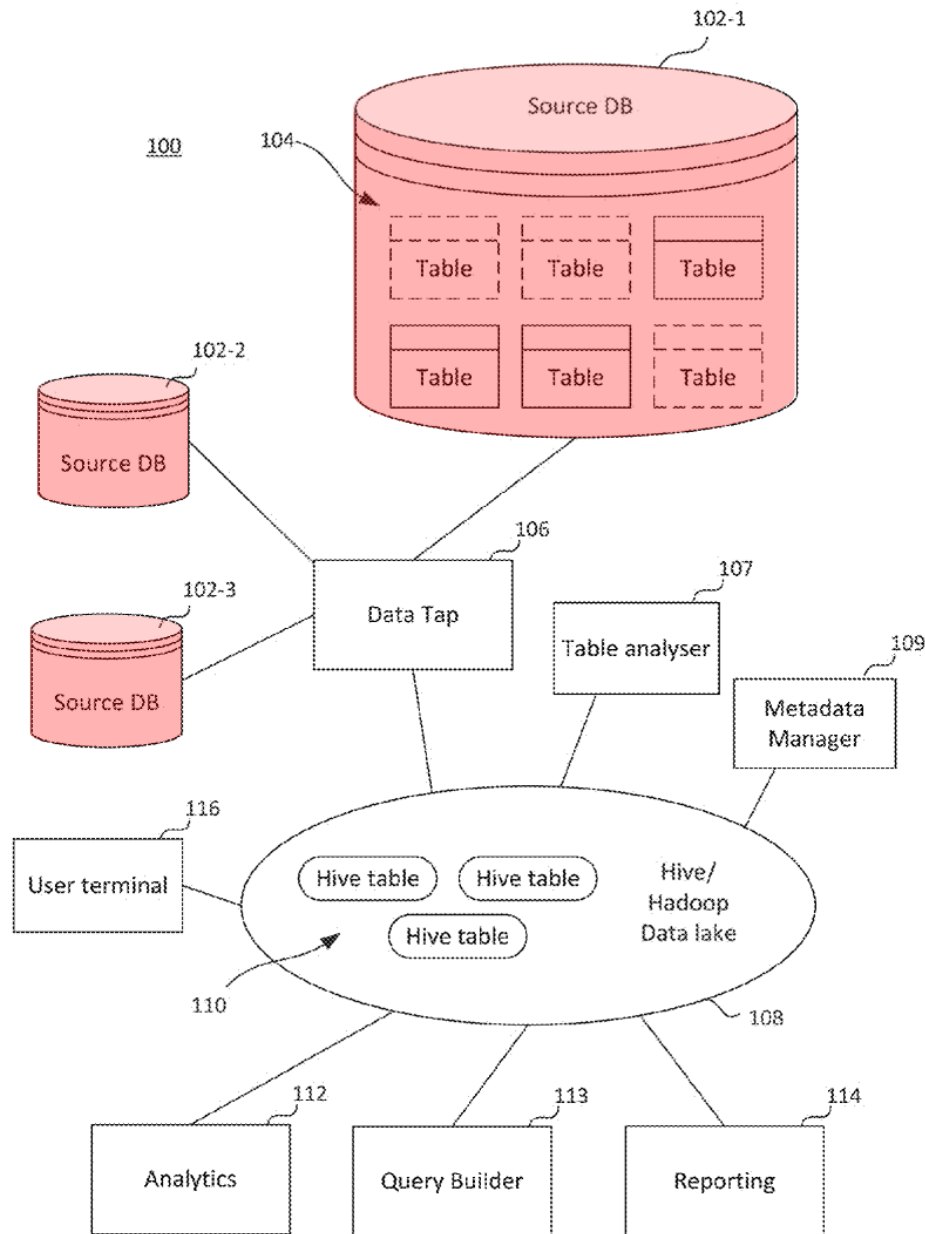


Fig. 1

EX1006, FIG. 1 (red annotations added).

79. Harrison further describes that “it may often be the case that data imported from disparate data sources is interrelated,” with an example of “a gas or similar utilities provider” that “import[s] a database of gas supply accounts from a

supply part of the organisation and a database of boiler maintenance data from a service/maintenance part of the organisation,” while “similar relationships may occur between disparate data sources maintained by organisations within any field (e.g. medical, banking, manufacturing etc.).” EX1006, [0065]. Harrison also contemplates heterogeneous source databases using different database management platforms. *Id.*, [0058]. Consistent with that distributed architecture, Harrison further discloses that its data-management server includes a “network interface . . . for connection to data sources 102” over a network such as the Internet. EX1006, [0456], Fig. 26. A POSITA would have understood from these disclosures that “disparate data sources” 102-1, 102-2, and 102-3 are implemented on different computer systems external to the local computer system at least because these data sources are separately maintained, accessible via a network, and comprise different databases. *Id.*, [0055], [0059], [0078]-[0090], [0454]-[0456], FIG. 26.

80. The combination of Harrison’s “disparate data sources” teachings with the event log generation framework of Rinke937 results in a local computer system creating an event log from raw data stored partly in one external computer system and partly in another. EX1004, [0012]-[0023], [0090], [0112]-[0113]; EX1006, [0055], [0059], [0064]-[0065], [0087], [0454]-[0457]. The ’708 patent’s Background likewise acknowledges that portions of a process may be executed and recorded on different computer systems. EX1001, 1:43-67. “[F]or example, the

ordering information for an order process and the history to it may be stored on a first computer system, while the data with respect to the sales and the vendor calls are stored in a second computer system.” *Id.*, 1:46-49.

81. **The Rinke937-Nooijen-Harrison computer system is a *local computer system*.** As noted above, the Rinke937-Nooijen-Harrison computer system comprises the “computer system having a processor and a storage means operatively coupled to the processor,” and it generates “an event log from raw data stored in a source system,” as taught by Rinke937. EX1004, [0012]; *id.*, [0013]-[0023] (describing the processor performing actions to generate the event log and storing the event log in the storage means), [0186] (identifying physical storage devices, including internal hard disks and flash memory); *see also id.*, [0051]-[0055], [0184]-[0188]. The computer system is a *local computer system* at least because it is “different” from the source system. *See* EX1001, 4:49-53 (“Hereby, the terms ‘external’ and ‘internal’ only indicate that different computer systems are concerned.”).

b) Element [1.1.b]

[1.1.b] wherein the raw data in the external computer systems is stored in a number of data tables,

82. As explained for element [1.1.a], Rinke937 teaches storing “raw data” in multiple data tables, and Rinke937-Nooijen-Harrison stores those multiple data

tables in external computer systems as taught by Harrison. EX1004, [0090]-[0102], [0124]-[0136]; EX1006, [0055], [0058]-[0059]; *supra* Section IX.E.1.a.

83. Rinke937 expressly states that “[i]n a source system which handles processes the data is usually saved into *tables* which reside in a data storage, e.g. in a database system.” EX1004, [0096] (emphasis added). Further, “[t]he different data fields which are sensed by the process sensor(s) are often scattered among different tables.” *Id.* Those fields contain the raw data used to create the event log: Rinke937’s event log generation process sensor units operate on “source data, i.e., raw data stored in the source systems” to identify process-step information. *Id.*, [0090]; *see also id.*, [0012]-[0023], [0075]-[0080], [0091]-[0102]. The process sensors units are defined using the source tables and fields containing the object identifiers, process-step information, timestamps, and additional attributes used to generate event-log records. *Id.*, [0099]-[0102]; *see also id.*, [0117]-[0118], Fig. 4.

84. Rinke937 provides a concrete example involving multiple source tables. Its “Change of content sensor” detects changes to an invoice using “the invoice table ‘BSEG’ and two change log tables ‘CDPOS’ and ‘CDHDR.’” EX1004, [0130]-[0135]. The unique identifier is retrieved from the invoice table BSEG, the process step information from the change log table CDPOS, and the timestamp from the change log table CDHDR; accordingly, “these three tables have to be related within this process sensor.” *Id.*, [0135]. Its invoice payment example likewise

retrieves the invoice identifier and payment-date information from the invoice table BSEG “in a standard SAP FI environment.” *Id.*, [0124]-[0129]. These disclosures expressly identify multiple source data tables holding the raw invoice and change information from which event-log records are generated. *Id.*, [0124]-[0136].

85. Thus, as Rinke937-Nooijen-Harrison includes the arrangement of multiple external computer systems as data sources taught by Harrison, Rinke937-Nooijen-Harrison stores the raw data used to create the process protocol in multiple data tables within the first and second external computer systems, thus teaching [1.1.b]. EX1004, [0090], [0096]-[0102], [0124]-[0136]; EX1006, [0055], [0058]-[0059].

c) Element [1.1.c]

[1.1.c] wherein the raw data comprises data, which is created during the execution of processes in the first external computer system and in the second external computer system,

86. Rinke937-Nooijen-Harrison satisfies this element. Rinke937 teaches that its raw source system data comprises records generated during process execution, and as explained above for element [1.1.a], Rinke937-Nooijen-Harrison includes multiple external source systems, as taught by Harrison, that provide such process-generated data from both the first and second external computer systems. EX1004, [0003]-[0008], [0075]-[0080], [0090]-[0096], [0124]-[0136]; EX1006, [0055], [0058]-[0065]; *supra* Section IX.E.1.a.

87. **Raw data created during the execution of processes:** Rinke937 explains that tasks reflecting process steps “are conducted in an IT driven environment and leave traces in an IT system,” and that “[p]icking up these traces and reconstructing the as-is process from this data is the goal of process mining.” EX1004, [0003]. Rinke937 accordingly teaches “deriv[ing] process data from the raw data comprising the digital traces of the executed processes.” *Id.*, [0080]; *id.*, [0008]. Those traces reside in the “raw data stored in the source systems.” *Id.*, [0090]; *id.*, [0103]. The process is reconstructed “solely from the traces which have been left on the source system.” *Id.*, [0075]. Rinke937 further explains that a process sensor unit derives process steps from the “raw data stored in the source systems” and that “[t]he respective changes which were triggered in the raw data are then converted into the aforementioned structure of an event log.” *Id.*, [0090]-[0091]; *see also id.*, [0072], [0076]-[0079], [0090]-[0096].

88. The concrete examples in Rinke937 confirm that these records arise during execution of operational processes. Its “Invoice paid sensor” senses “when an invoice has been paid in an ERP-system,” using the invoice’s “clear date” as “the timestamp of the process step.” *Id.*, [0124]-[0129]. Its “Change of content sensor” likewise senses process steps involving changes to invoice contents, using fields recorded by the source system in change logs and the corresponding “change date.” *Id.*, [0130]-[0134]. The invoice identifier, process step, and timestamp are retrieved

from the invoice table BSEG and change-log tables CDPOS and CDHDR. *Id.*, [0135]. These disclosures identify records created or updated in the “source systems” as the underlying processes occur and subsequently used to generate the event log. *Id.*, [0090], [0103], [0124]-[0136].

89. **Process-generated data in the first and second external computer systems:** Harrison teaches consolidating data from “many disparate databases,” including databases that “support[] different operations of the organisation,” because “it may often be the case that data imported from disparate data sources is interrelated.” EX1006, [0064]-[0065]. One example includes “a database of gas supply accounts from a supply part of the organisation” and “a database of boiler maintenance data from a service/maintenance part of the organisation.” *Id.*, [0065]. These databases contain related records because “some supply customers may also be maintenance customers.” *Id.* A POSITA would have understood this example as teaching related processes executing on multiple external computer systems, including at least a first external computer system and a second external computer system, because it was well known that raw process data was often generated and stored across multiple different source systems. *Supra* Section VI.B.2. Harrison also expressly contemplates changing operational records, preserving “older versions of records which have since been updated” and performing subsequent imports “to

capture changes in the data sources and apply those changes to the data lake.”
EX1006, [0063], [0205].

90. Thus, in Rinke937-Nooijen-Harrison, the raw data used to create the event log includes data generated during execution of processes in each of the first and second external computer systems. EX1004, [0075]-[0080], [0090]-[0096], [0124]-[0136]; EX1006, [0064]-[0065].

d) Element [1.1.d]

[1.1.d] wherein a process comprises several process steps, wherein each process step creates the data during execution, and wherein the method comprises:

91. Rinke937-Nooijen-Harrison satisfies this element. In Rinke937-Nooijen-Harrison, a process comprises multiple process steps, whose execution creates source system data, which in turn is used to generate the event log, as taught by Rinke937. EX1004, [0003]-[0008], [0075]-[0080], [0090]-[0096], [0120], [0124]-[0136].

92. **A process comprises several process steps:** Rinke937 expressly states that “[b]usiness processes consist of multiple process steps.” EX1004, [0120]. It describes those steps as activities “conducted with the process element,” such as “Invoice is received” or “Registration of the patient in the emergency room,” and explains that the steps have an order; for example, a patient is treated before a bill is sent to the insurance company. *Id.*, [0004]-[0007]. Rinke937 accordingly identifies

“[a] set of process steps which a process element can pass during its process” and “[t]he order of the process steps.” *Id.*, [0077]-[0078]. Its exemplary event log records successive steps including “Purchase Requisition,” “Purchase Requisition Approval,” “Purchase Order Item,” “Goods Receipt,” and “Invoice Receipt,” together with their corresponding timestamps. *Id.*, [0086]; *see also id.*, [0081]-[0088], [0112]-[0113], [0120], Figs. 2, 5. These disclosures expressly identify processes comprising several steps.

93. **Each process step creates data during execution:** Rinke⁹³⁷ explains that “[m]ost tasks which are reflecting steps on the way of the process are conducted in an IT driven environment and leave traces in an IT system.” EX1004, [0003]. Its “process sensor” executables derive process data from “raw data comprising the digital traces of the executed processes,” permitting reconstruction of the process “solely from the traces which have been left on the source system.” *Id.*, [0075], [0080]; *see also id.*, [0008]. After sensing “one or multiple process steps” in the raw data, “[t]he respective changes which were triggered in the raw data are then converted into the aforementioned structure of an event log.” *Id.*, [0091]. The sensed data includes process step descriptions, corresponding timestamps or time spans, and the identifier of the process element. *Id.*, [0092]-[0094]; *see also id.*, [0076]-[0079], [0090], [0096]-[0102].

94. Rinke937's examples illustrate the data created by individual process steps. Its "Invoice paid sensor" detects "when an invoice has been paid in an ERP-system" based on the invoice having "a valid date in the field 'clear date.'" *Id.*, [0124]. The sensor uses that field's contents as the process-step timestamp and identifies the step as "Invoice paid." *Id.*, [0125]-[0129]. Likewise, its "Change of content sensor" detects "all process steps which belong to a change of the content of an invoice" using fields recorded by the source system in change logs. *Id.*, [0130]. The resulting event identifies the changed field—for example, "Change of: invoice amount"—and uses the field's "change date" as the timestamp. *Id.*, [0131]-[0133]. Rinke937 identifies the underlying data sources as invoice table BSEG and change log tables CDPOS and CDHDR. *Id.*, [0134]-[0136].

95. A POSITA would therefore have understood that the recorded process steps in these disclosed processes create corresponding source system raw data during execution, including payment date entries and records of invoice changes. The combination subsequently uses that raw data to generate the event log. In Rinke937-Nooijen-Harrison, a process comprises several process steps, each creating the data during execution, thus teaching [1.1.d]. EX1004, [0003]-[0008], [0075]-[0080], [0090]-[0096], [0120], [0124]-[0136].

2. **Element [1.2]**

[1.2] (a) importing the data tables from the first external computer system and from the second external computer system into the local computer system, and storing the imported data tables in the storing device of the local computer system,

96. Rinke937-Nooijen-Harrison satisfies this element. As discussed above, Rinke937-Nooijen-Harrison *imports the data tables from the first external computer system and from the second external computer system into the local computer system*. *Supra* Section IX.D. Rinke937-Nooijen-Harrison “import[s] data from any number of data sources 102-1, 102-2, 102-3 into a data repository 108,” as taught by Harrison. EX1006, [0055], [0059], Fig. 1; *supra* Sections IX.D (Rinke937-Nooijen-Harrison combination), IX.E.1.a (element [1.1.a]). Those data sources include separate relational databases, with each source database containing multiple data tables. *Id.*, [0058]-[0059]; *see also id.*, [0003] (“Often there is a need to bring together data from diverse data sources to enable processing and analysis of data sets.”). Rinke937-Nooijen-Harrison imports “a database in its entirety (i.e. including all tables)” or “may import only one or more selected tables” from the databases stored in respective data source systems, as taught by Harrison. *Id.*, [0059]; *see also id.*, [0061] (explaining that “all columns of the source table will be imported”), [0089] (disclosing importing “the full table contents”), [0159]-[0164] (describing

process for importing the “full table data set” into the corresponding centrally stored table).

97. Rinke937-Nooijen-Harrison stores the imported table data in the storage device of its computer system (*storing the imported data tables in the storing device of the local computer system*) so that the computer system can “analyse and process the data centrally,” as taught by Harrison. EX1006, [0064]. As explained above, Rinke937-Nooijen-Harrison combines the table-import functionality of Harrison with the computer system (*local computer system*) taught by Rinke937, which includes “a storage means being operatively coupled to the processor.” EX1004, [0052]-[0055]; *supra* Section IX.D. Harrison teaches importing source tables into a data repository that may use “any data storage technology,” including “a conventional structured database such as a relational database.” EX1006, [0055]-[0056], [0059]. Harrison also expressly permits the computer system storing the imported data to “combine data storage with any of the previously described processing functions.” *Id.*, [0461]. Its teachings thus support maintaining the imported table copies within the same computer system that processes them.

98. Rinke937-Nooijen-Harrison thus stores the imported table copies taught by Harrison in the local storage taught by Rinke937, for analysis using the automated techniques taught by Nooijen. *Supra* Section IX.D. This arrangement makes the table contents locally available for repeated examination and comparison

when reconstructing table relationships, without repeatedly retrieving those contents from the external source systems. *See* EX1005, 319-20, 324-25. The imported tables therefore reside in the storage device of the local computer system, while the original tables remain in the external computer systems.

3. Element [1.3]

[1.3] (b) in the local computer system, determining and reconstructing the relationships between the imported data tables,

99. Rinke937-Nooijen-Harrison satisfies this element, including under the proposed construction where relationships between tables imported from different computer systems were unknown before the determining and reconstructing. *Supra* Section VIII. Nooijen expressly teaches that “first the schema SR including column types and primary and foreign keys is rediscovered,” including “non-documented foreign key relations.” EX1005, 317. Nooijen explains that schema extraction techniques “rediscover foreign keys between tables” and identify “undocumented foreign key relations (used in the application, but not declared as such).” *Id.*, 319. The result is a relational schema containing the table schemas and the identified foreign keys. *Id.* Nooijen thus determines relationships from the table data and reconstructs a schema representing those relationships.

100. Figure 1 of Nooijen depicts a “Structured dataset” as input to the discovery workflow, which performs “Schema extraction” to produce a “Database schema” and then identifies “artifact schema(s)”:

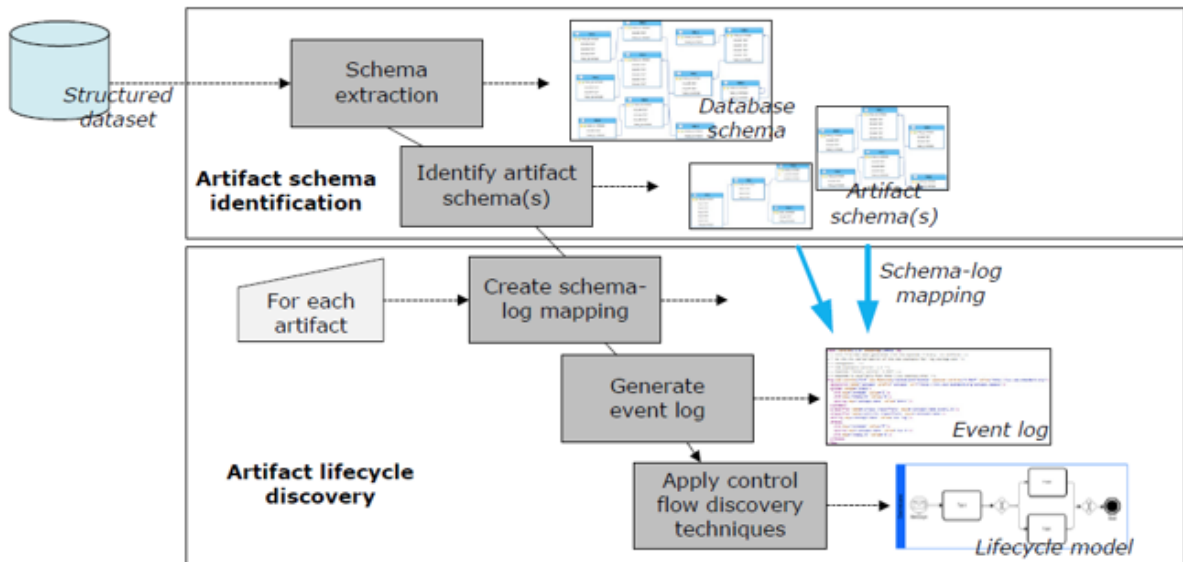


Fig. 1. Approach to Artifact Discovery

EX1005, Fig. 1; *id.*, 317.

101. The “foreign key relations” in Nooijen are relationships between data tables. Nooijen describes each foreign key as connecting a parent table and its primary-key columns with a child table and its referencing columns. EX1005, 319. Identifying that connection determines which columns relate records in one table to records in another, and including the connection in the reconstructed schema represents the relationship between those tables. *Id.* For example, identifying that a column in an activity table references the primary-key column in an object table

establishes how the activity records are associated with particular objects. The reconstructed schema records the participating tables and columns so that subsequent operations can use those relationships to join their records. This is consistent with the description in the '708 patent, which identifies the reconstructed relationships between imported tables as “foreign key relationships.” EX1001, 12:10-13. Harrison uses the same terminology, describing its analysis of imported tables as identifying “a primary key to foreign key relationship,” meaning a relationship between a primary key in one table and a corresponding foreign key in another. EX1006, [0221].

102. Rinke937-Nooijen-Harrison performs this determining and reconstructing, as taught by Nooijen, on the imported tables stored in the local computer system, as taught by Harrison. As explained above, Rinke937-Nooijen-Harrison uses the import functionality in Harrison to bring source tables into the local storage in Rinke937 for automated analysis. *Supra* Sections IX.D, IX.E.2. Nooijen calls its input a “relational database R,” such as “an ERP system’s database.” EX1005, 319. That input corresponds to the relational repository containing the imported copies of the raw source-system tables described in Rinke937 and Harrison. EX1004, [0090], [0096]; EX1006, [0056], [0059]; *supra* Sections IX.E.1.b, IX.E.2. The “schema” reconstructed in Nooijen describes the

columns, keys, and determined relationships of those locally stored tables, including “undocumented foreign key relations” between the tables. EX1005, 319.

103. Harrison further confirms the suitability of the imported copies for that analysis: its Table Analyser examines “the data contents of tables imported into the data lake” to identify relationships between them. EX1006, [0221], [0224]-[0225]; *see also id.*, [0067]-[0068], [0222]-[0223], [0226]-[0235]. Harrison expressly contemplates relationships between data from different source systems, including overlapping customer identifiers in separately maintained databases. EX1006, [0065]. It explains that documentation may be “incomplete or inconsistent” and that existing table relationships may be lost during import. *Id.*, [0066]. Its relationship-discovery techniques address those problems, identifying relationships that permit joins between source tables and using the identified relationships to create join queries. *Id.*, [0067]-[0068]. These teachings reinforce applying the automated schema reconstruction in Nooijen to the combined collection of locally imported tables, including tables originating in different external systems. *Supra* Sections IX.D, IX.E.2.

104. The cross-system relationships identified through this analysis include relationships that were previously unknown. Harrison explains that when “tables ... originated from different data sources ... the relationships between the tables may not be known a priori.” EX1006, [0421]. Harrison describes importing a database of

gas-supply accounts from one part of an organization and a database of boiler-maintenance data from another part of the organization. EX1006, [0065]. Some customers appear in both databases, creating relationships through overlapping customer information. *Id.* A POSITA would have understood that each source application could operate using its own customer records without identifying the columns connecting those records to customer records in the other database. Knowledge of relationships within either source database therefore does not establish knowledge of the relationships between the two databases. As Harrison explains, related data from different sources need not even reside in tables or columns with matching or related names. *Id.*, [0066]. Identifying the cross-system relationships consequently involves analysis of the combined data after import.

105. This was a recognized reason for automated foreign-key discovery. Zhang, a background reference cited by Nooijen, explains that database designers may omit foreign-key constraints because “they are not aware of implicit relationships inherent in the data” and because “such relationships might hold across multiple databases.” EX1009, 805; *see* EX1005, 319 (citing Zhang, reference [21], for its techniques “to rediscover foreign keys between tables”), 327 (reference [21]). Zhang describes discovery over tables potentially originating in multiple databases without requiring prior knowledge of the data. EX1009, 806. A POSITA therefore

would have understood that these techniques could discover previously unidentified relationships connecting different databases.

106. The schema-discovery techniques taught by Nooijen determine relationships from the table contents without requiring those relationships to be specified beforehand. Nooijen explains that these techniques identify “undocumented foreign key relations” and represent the discovered connections in a reconstructed schema. EX1005, 319. A POSITA would have understood that a relationship can exist in the data before it has been identified: corresponding values in different tables may associate records with the same object, while the particular tables and columns establishing that association remain unknown until the data is analyzed. In Rinke937-Nooijen-Harrison, the discovery techniques operate on the imported table contents to identify those connections across source systems. Harrison expressly describes relationships between tables from different data sources that “may not be known a priori,” and identifies them through analysis of the imported contents. EX1006, [0065]-[0068], [0421]. Zhang likewise confirms that foreign-key discovery addresses implicit relationships that are unknown, including relationships spanning different databases. EX1009, 805-806. These teachings support applying the automated discovery in Nooijen to determine previously unknown cross-system relationships and represent them in the reconstructed schema.

107. A POSITA would have applied the automated schema-discovery techniques taught by Nooijen to the combined collection of imported tables in those circumstances. Doing so would identify connections needed to associate related process data from different systems without requiring a predefined cross-system mapping. *Supra* Section IX.D. The analysis evaluates candidate connections between columns using the imported data, determines which columns connect the related records, and represents the identified relationships in the reconstructed schema. EX1005, 319; EX1006, [0068], [0221]-[0225]. Thus, in Rinke937-Nooijen-Harrison, previously unknown cross-system relationships are determined after import through analysis of the table contents and then represented in the reconstructed schema.

108. This reconstruction supplies the table relationships used in creating the executable process sensor units taught by Rinke937. Rinke937 explains that relevant fields are “scattered among different tables” and that “[t]he relations between the different tables have to be defined within the process sensor(s).” EX1004, [0096]. Those relationships permit joins between tables containing object identifiers, process-step information, and ordering attributes. *Id.*, [0099]-[0102]. Rinke937 illustrates this with an executable process sensor unit that retrieves the invoice identifier from BSEG, process-step information from CDPOS, and the timestamp from CDHDR, explaining that “these three tables have to be related within this

process sensor.” EX1004, [0135]. In the combination, the foreign-key relationships reconstructed through Nooijen identify the columns connecting those related tables and supply the relationships needed to define the sensor joins. EX1005, 319; *supra* Section IX.D.

109. Accordingly, the Rinke937-Nooijen-Harrison local computer system determines and reconstructs relationships between the imported data tables after import, including previously unknown relationships between tables originating in different external computer systems, thus teaching element [1.3].

4. **Element [1.4]**

[1.4] (c) in the local computer system, determining and selecting a first data table from the number of imported data tables, wherein each data record stored in the first data table represents an object for which the process protocol is to be created,

110. Rinke937-Nooijen-Harrison satisfies this element. Nooijen identifies “a unique main table . . . with the least distance to all other tables in the cluster” and designates that table as the main table of an artifact schema. EX1005, 320. In Rinke937-Nooijen-Harrison, this identification and selection operates on the locally imported tables described for elements [1.2] and [1.3]. *Supra* Sections IX.D, IX.E.2-IX.E.3.

111. **Rinke937-Nooijen-Harrison determines and selects a first data table.** The “main table” in Nooijen is a table identifying the business objects whose activities are to be recorded in the event log. EX1005, 316, 319-320. Nooijen

explains that “[t]he idea of artifacts is to decompose this data into the business objects, or artifacts, of the process.” EX1005, 319. An artifact schema describes the information of one artifact and “consists of all tables that contain relevant information” about that artifact. *Id.*, 317. Nooijen identifies these schemas “automatically through clustering,” grouping tables based on their contents and foreign-key relationships. *Id.*, 320. Within each cluster, Nooijen identifies the main table and distinguishes it from the other tables describing additional attributes of the artifact. *Id.* Identifying and designating that main table is *determining and selecting a first data table from the number of imported data tables*.

112. **Each record represents an object for which the process protocol is to be created.** Nooijen explains that each “artifact instance has a unique identifier and follows a life-cycle describing how attributes of the artifact change as the process evolves,” and that “the primary key of the main table is the identifier of the artifact.” EX1005, 320. A POSITA would therefore have understood each record of the main table, distinguished by its primary-key value, to represent the respective artifact instance (*object*). This main table corresponds to the first or base table described in the ’708 patent, which identifies business-object tables containing base information about orders or vendors and designates the selected business-object table as the first data table. EX1001, 13:38-54, 13:64-67.

113. Nooijen expressly connects those object instances to the event log to be created. Its technique extracts a log describing the instances of the artifact “and their evolution over time.” EX1005, 317. During extraction, event information is joined with the main table, “thus associating each event with exactly one trace.” *Id.*, 324. The events are then grouped by the trace-identifier values and ordered by timestamp; “[e]ach group defines a trace,” and the resulting traces are written into the event log. *Id.* These identifiers correspond to the process-object identifiers used in Rinke937 to associate process steps with individual objects, as explained for element [1.1.a]. EX1004, [0081]-[0084], [0099]; *supra* Section IX.E.1.a. Each record of the main table therefore represents an object whose associated activities populate the resulting event log (*process protocol*).

114. As explained for elements [1.2] and [1.3], the Rinke937-Nooijen-Harrison computer system stores the imported source tables and reconstructs their schema using the techniques in Nooijen. *Supra* Sections IX.E.2-IX.E.3. Applying the main-table identification in Nooijen to that reconstructed schema selects a table from those locally imported tables. The selected table supplies the object identifiers used to configure the process sensor executable units taught by Rinke937. *Supra* Section IX.D. Accordingly, Rinke937-Nooijen-Harrison teaches element [1.4].

5. **Element [1.5]**

[1.5] (d) in the local computer system, determining and selecting a number of second data tables from the number of imported data tables, which are necessary for the creation of the process protocol, based on the selected first data table and based on the reconstructed relationships between the imported data tables, wherein each data record stored in the number of second data tables represents a process step,

115. Rinke937-Nooijen-Harrison satisfies this element. As explained for element [1.4], Rinke937-Nooijen-Harrison selects a main table whose records identify the objects for which the event log is to be created, as taught by Nooijen. EX1005, 321-24; *supra* Section IX.E.4. Rinke937-Nooijen-Harrison then applies the automated mapping techniques taught by Nooijen to identify related event tables (*second data tables*) supplying the process-step information for those objects, using the reconstructed schema and foreign-key relationships. EX1005, 321-24.

116. **Rinke937-Nooijen-Harrison determines and selects second data tables based on the selected first data table and the reconstructed relationships.** Nooijen identifies schema-to-log mappings through an automated approach “based on timestamps and foreign key relations,” using the relationships reconstructed in the preceding schema-discovery stage. EX1005, 321; *see also id.*, 317, 319-20; *supra* Section IX.E.3. Nooijen identifies event types by “selecting all columns with a timestamp domain as event columns, except for columns that are part of a parent table of the main table.” EX1005, 322. For each selected event column, Nooijen

identifies the corresponding event table containing that column. *Id.*, 322-23. Thus, the selection expressly depends on both the selected main table and the reconstructed parent-child relationships between tables. The selected event tables other than the main table are *second data tables*.

117. Nooijen then links each event table to the selected main table. Algorithm 1 passes the main table and the event table to the CreateMapping algorithm, which determines the path between them. EX1005, 322 (Algorithms 1, 2). Nooijen explains that this path “consists of the sequence of foreign keys connecting those tables.” *Id.*, 323. These links associate the process-step information in the second tables with the objects identified by the first table. *Id.* They also supply the table relationships used to define joins in the process sensor units taught by Rinke937. EX1004, [0096]-[0102]; *supra* Section IX.D.

118. **Rinke937-Nooijen-Harrison selects second data tables whose records represent process steps.** Rinke937 provides a concrete example in which its “Change of content sensor” senses process steps involving changes to invoice contents using change-log tables CDPOS and CDHDR, which supply the changed-field information and corresponding change date. EX1004, [0130]-[0135]. In Rinke937-Nooijen-Harrison, the automated mapping techniques taught by Nooijen are applied to such event tables, whose records document changes or actions performed in the source system. Nooijen identifies event types through timestamp

columns and explains that “[e]xactly one event should be created for each value” in an identified event column. EX1005, 321; *see also id.*, 322-23. For such tables, with a dedicated time-stamp column (e.g., C_e in Nooijen), the timestamp specifies when the recorded step occurred, while the primary key of the event table identifies the particular record used in the event mapping. *Id.*, 322-23. Thus, each record in the selected event table represents a recorded process step, and the mapping identifies how to extract that occurrence as an event. Joining those records with the main table adds the corresponding object identifier, “thus associating each event with exactly one trace.” EX1005, 324. Accordingly, applying the teachings of Nooijen to the process-step records taught by Rinke937, Rinke937-Nooijen-Harrison selects event tables (*second data tables*) in which each stored record represents a process step.

119. **Rinke937-Nooijen-Harrison selects second data tables necessary to create the process protocol.** The selected event tables (*second data tables*) in Nooijen supply the event identifiers and timestamps needed to record the activities of the selected objects. Nooijen extracts those values by joining the event tables with the main table through the identified foreign-key relationships. EX1005, 324. Nooijen explains that “[e]ach record in the joined table defines an event with the given name” and that the joined table contains the main table’s trace-identifier columns, “thus associating each event with exactly one trace.” *Id.* Nooijen then groups the events by object identifier and orders them by timestamp to create the

event log. *Id.* The selected event tables thus provide the process-step information needed to populate the process protocol for the objects identified by the selected first table.

120. **Rinke937-Nooijen-Harrison performs this selection in the local computer system from the imported data tables.** As explained for elements [1.2]-[1.4], the Rinke937-Nooijen-Harrison computer system imports and locally stores the source tables, reconstructs their relationships, and selects the main table. *Supra* Sections IX.E.2-IX.E.4. Applying the event-table selection in Nooijen to those locally stored tables and their reconstructed schema determines and selects the second data tables within that same system. *Supra* Section IX.D. Accordingly, Rinke937-Nooijen-Harrison teaches element [1.5].

6. Element [1.6]

[1.6] (e) in the local computer system, determining activities for the objects of the selected first data table based on the selected second data tables and the reconstructed relationships between the selected first data table and the selected second data tables,

121. Rinke937-Nooijen-Harrison satisfies this element. As explained in the following paragraphs, Rinke937-Nooijen-Harrison applies the automated event-identification and mapping techniques taught by Nooijen to determine activities for the objects identified by the main table (*selected first data table*), using the event

tables (*selected second data tables*) and the reconstructed relationships. *Supra* Sections IX.D, IX.E.4, IX.E.5.

122. **Rinke937-Nooijen-Harrison determines activities for the objects of the selected first data table.** Nooijen explains that “all event types in the artifact schema are identified automatically.” EX1005, 321. Its schema-log mapping defines those event types and maps the tables and attributes of the artifact to the event types. *Id.* For each event type, Algorithm 1 creates an event mapping containing an event name, event-identifier columns, timestamp column, and links to the relevant tables. *Id.*, 321-22. Nooijen further explains that “the event column C_e and an event name are added” to create the event mapping. *Id.*, 323. These named event types identify the activities to be recorded, while the event identifiers and timestamps identify particular occurrences of those activities. This corresponds to the process-step descriptions and ordering attributes used by Rinke937 to generate entries for the event log, including named activities such as “Invoice paid.” EX1004, [0092]-[0094], [0100]-[0101], [0124]-[0129].

123. **Rinke937-Nooijen-Harrison determines those activities based on the selected second data tables and reconstructed relationships.** Nooijen identifies event information through an approach “based on timestamps and foreign key relations,” using the relationships obtained through its preceding schema-discovery techniques. EX1005, 321; *see also id.*, 322-23. Algorithm 1 constructs

each event mapping using the selected main table and the corresponding event table, and Algorithm 2 then determines the path connecting those tables. EX1005, 322. This path “consists of the sequence of foreign keys connecting those tables.” *Id.*, 323. In Rinke937-Nooijen-Harrison, the event columns in the event tables (*selected second data tables*) supply the activity information, and the relationships reconstructed for element [1.3] connect that information to the objects represented in the main table (*selected first data table*). *Supra* Sections IX.E.3-IX.E.5.

124. **Rinke937-Nooijen-Harrison associates the determined activities with the respective objects represented in the first data table.** The Nooijen event mappings identify how to join the tables containing the event identifier and timestamp with the main table, including intermediate tables where required by the foreign-key relationships. EX1005, 324. Nooijen explains that “[e]ach record in the joined table defines an event with the given name,” and contains the trace-identifier columns from the main table, “thus associating each event with exactly one trace.” *Id.* As explained for element [1.4], those trace identifiers identify the individual objects represented by the records of the main table. *Supra* Section IX.E.4. The mappings therefore determine which activities relate to each identified object, supplying the object identifiers, process-step descriptions, and ordering information used by the process sensors in Rinke937. EX1004, [0099]-[0102]; *supra* Section IX.D.

125. Rinke937-Nooijen-Harrison performs this activity determination in the local computer system. As explained for the preceding elements, the Rinke937-Nooijen-Harrison computer system imports and locally stores the source tables, reconstructs their relationships, and selects the first and second tables. *Supra* Sections IX.E.2-IX.E.5. It then applies the event-identification and mapping techniques in Nooijen to those locally stored tables and their reconstructed relationships, consistent with the combination described above. *Supra* Section IX.D. Accordingly, Rinke937-Nooijen-Harrison teaches element [1.6].

7. **Element [1.7]**

[1.7] (f) in the local computer system, creating an executable transformation, which based on

[i] the selected first data table,

[ii] the selected second data tables, and

[iii] the relationships between the selected first data table and the selected second data tables

[iv] is adapted to extract data from the first data table and from the second data tables based on the determined activities, and to store the extracted data as process protocol, and

126. Rinke937-Nooijen-Harrison satisfies this element.

127. **Rinke937-Nooijen-Harrison creates a process sensor (*executable transformation*) in its computer system (*local computer system*).** Rinke937 teaches “creation of the process sensors” using “relations, source tables, target tables

and other attributes.” EX1004, [0116]-[0119]. Rinke937 repeatedly describes its process sensors as “executable code.” *E.g., id.*, [0119], [0012], [0027], [0075], [0098], [0110]-[0114] (passage titled “Execution of One or Multiple Process Sensors”), claims 1, 4, 7-10.

128. Rinke937 further discloses that these executable process sensors carry out transformations “to transform the source data into an event log format for further process mining analysis.” *Id.*, [0010]; *id.*, [0072] (“This transformation is conducted by process sensors, ...”), [0096]-[0102] (describing process sensors and associated transformations used to generate event logs, including “transformation of the fields ... over (possibly) multiple tables” and “transformation using the fields from all joined tables”). Rinke937’s event log generation framework creates “the executable code for the underlying execution platform” using the “relations, source tables, target tables and other attributes” through either a scripting language or a graphical user interface. *Id.*, [0117]-[0119]; *see also id.*, [0115]-[0116]. Rinke937 expressly discloses converting a process sensor statement from an independent representation into “an executable representation for being executed with a predetermined execution environment.” EX1004, [0031]-[0032]. The resulting executable sensor code implements the process sensor units that extract process data from the source systems and create the event log data structure. *Id.*, [0090]-[0091], [0113], [0120]-[0121]; *see also id.*, [0027], [0056]-[0063]. POSITAs would thus understand that the

Rinke937 process sensors are “executable transformations” as claimed by the ’708 patent.

129. Rinke937 also teaches creating its process sensors on the local computer system, i.e., the Rinke937 “computer system.” For example, Rinke937 discloses that its “process sensor may be executed by executing at least one sensor statement of the at least one process sensor” and that “the processor converts the independent representation of the sensor statement into an executable representation.” EX1004, [0031]; *id.*, claim 8 (“the processor converts the independent representation of the sensor statement into an executable representation for being executed”), claim 1 (“a computer system having a processor”). It further teaches generating “executable code” from the scripted process sensor statements “during runtime.” Rinke937, [0119] (“[F]rom the scripted programming language the executable code ... is generated during runtime.”), [0117] (“In the scripted interface [for process sensor creation,] the relations, source tables, target tables and other attributes are scripted in a programming language.”). Given these teachings, a POSITA would have naturally implemented the combined system such that the process sensors (*executable transformations*) taught by Rinke937 were generated by the computer system (*local computer*). *Supra* Section IX.D, IX.E.2-IX.E.6 (describing other “local computer system” elements); EX1001, 4:49-53 (“[T]he

terms ‘external’ and ‘internal’ only indicate that different computer systems are concerned.”).

130. Nooijen confirms the suitability of its mappings for generating executable extraction code. Nooijen explains that its Algorithm 1 supplies the event-log mapping “fully automatically” and that, upon receiving that mapping, “XESame automatically generates the database queries needed to extract the log.” EX1005, 324. This implemented mapping-to-query workflow supports using the same automatically determined table, field, and relationship information to supply sensor definitions within the code-generation framework in Rinke937. *Supra* Section IX.D.

131. **Rinke937-Nooijen-Harrison creates the process sensor (*executable transformation*) based on the main table (*selected first table*), the event tables (*selected second tables*), and the foreign-key relationships between them (*relationships between the selected first data table and the selected second data tables*) (sub-elements [i]-[iii]).** In Rinke937-Nooijen-Harrison, the automated mappings from Nooijen supply the information used to define the process sensors: the main-table identifier columns identify the fields used to derive the “unique identifier” corresponding to the “process object”; the event mappings specify the “process step description” and the “order attribute,” such as a timestamp; and the reconstructed foreign-key relationships define the “consecutive join over (possibly) multiple tables” used to retrieve that information. EX1004, [0099]-[0101]; EX1005,

321-24; EX1004, [0117] (noting “the relations, source tables, target tables and other attributes” of a process sensor script), [0096]-[0102] (describing those attributes in more detail). A POSITA would have implemented Rinke937-Nooijen-Harrison to use Nooijen’s main table (*selected first data table*), related event table (*selected second data table*), and foreign-key relations connecting them (*relationships between the selected first and second data tables*) to create the scripts from which the process sensor executables are generated, as taught by Rinke937. EX1005, 321-23; EX1004, [0096]-[0102], [0117]-[0119]; *supra* Sections IX.E.3-IX.E.5.

132. Nooijen teaches that Algorithm 1 passes the main table and each selected event table to CreateMapping; Algorithm 2 determines the connecting path, which “consists of the sequence of foreign keys connecting those tables.” EX1005, 322-23. Those results supply the source tables, identifier fields, and joins used to define the process sensor units in Rinke937. EX1004, [0099]-[0102], [0117]. Because Rinke937 generates executable code from those definitions, the resulting transformation is based on the selected first data table, the selected second data tables, and their reconstructed relationships. EX1004, [0119].

133. To illustrate, suppose the selected main table identifies orders through an OrderID column and a related event table records shipment occurrences in a ShipmentDate column. Applying Nooijen, the generated mapping identifies ShipmentDate as the event column, uses its column name as the event name, and

records the foreign-key path connecting the event table to the main table. EX1005, 321-323. Those results supply concrete sensor-definition inputs: the OrderID field of the main table supplies the process-object identifier, the fixed label “ShipmentDate” supplies the process-step description, the values read from ShipmentDate supply the timestamp, and the foreign-key path supplies the joins associating each shipment event with its order. EX1004, [0099]-[0102]. Rinke937-Nooijen-Harrison then generates executable code from those table, field, and relationship definitions, implementing the automatically determined mapping as an executable process sensor unit as taught by Rinke937. EX1004, [0117]-[0119].

134. **Rinke937-Nooijen-Harrison process sensors (*executable transformations*) extract data from the selected tables based on the determined activities and store that data as an event log (*process protocol*) (sub-element [iv]).** Rinke937 teaches that its process sensor executable units extract “process data” from the identified tables and store that data in an “event log” (*process protocol*) *E.g.*, EX1004, [0090] (“A process sensor is able to sense the necessary data for a process step ...”), [0080] (“[A] process sensor is adapted to derive process data ...”), [0096] (“The different data fields which are sensed by the process sensor(s) are often scattered among different tables.”), [0091] (describing how process sensors “sense one or multiple process steps” and store “respective changes” in an event log); *see also id.*, [0092]-[0094], [0096]-[0102], [0120]-[0121]; *supra* Section IX.E.1.a

(element [1.1.a]). The process sensors generate various element/object identifiers, process-step identifiers, and order-of-process information, then store “the generated data as event log with the storage means.” EX1004, [0013]-[0023].

135. The extracted data stored in the event log (*process protocol*) is based on the determined activities. In the combination, the Nooijen event mappings supply the object and activity information used to populate that event log, while Rinke937 supplies the executable event log generation framework and storage operation. EX1004, [0020]-[0023], [0097]-[0102], [0119]-[0121]; EX1005, 321-24. Each Nooijen event mapping specifies an event name (*determined activity*), event identifier, timestamp, and table relationships identifying how to retrieve the associated information. EX1005, 321-23; *supra* Section IX.E.6. The extraction defined by that mapping joins the tables containing the event identifier and timestamp with the main table, using the identified foreign-key relationships. EX1005, 324. The resulting records contain the object identifier from the main table (*selected first data table*) and event information from the event tables (*selected second data tables*). *Id.* Nooijen explains that “[e]ach record in the joined table defines an event with the given name” and that the main table’s trace-identifier columns associate “each event with exactly one trace.” *Id.*

136. In Rinke937-Nooijen-Harrison, this information defines the process-object identifier, process-step description, ordering attribute, and joins in Rinke937.

EX1004, [0099]-[0102]. The generated process sensors therefore implement the extraction specified by the event mappings: they obtain the object identifiers and associated process step data from the selected tables according to the determined activities. EX1004, [0090]-[0102], [0119]-[0121]. Extraction from the original tables in the external computer systems is further explained for element [1.8]. *Infra* Section IX.E.8.

137. Accordingly, Rinke937-Nooijen-Harrison satisfies element [1.7].

8. Element [1.8]

[1.8] (g) in the local computer system, executing the transformation, wherein during execution of the transformation the data is extracted from the data tables stored in the external computer systems (ES1, ES2), and storing the extracted data as process protocol, and providing the created process protocol to a Process Mining System for analysis and reconstruction of the processes carried out in the first external computer system and in the second external computer system.

138. Rinke937-Nooijen-Harrison satisfies this element. As explained for element [1.7], the Rinke937-Nooijen-Harrison computer system generates executable process sensor code (*the transformation*) using the tables, fields, activities, and relationships determined through the teachings in Nooijen. *Supra* Sections IX.D, IX.E.7. The imported table copies supply the data used to develop those sensor definitions. *Supra* Sections IX.E.2-IX.E.7. When Rinke937-Nooijen-Harrison executes the process sensor code, the system reads the raw data in the corresponding original tables in the external source systems to obtain current process

data, stores the resulting event log (*process protocol*), and provides the log to a process mining system, as taught by Rinke937. EX1004, [0012]-[0023], [0035], [0090]-[0102], [0112]-[0113]; EX1006, [0087]-[0090], [0456].

139. **The Rinke937-Nooijen-Harrison computer system (*local computer system*) executes the process sensor (*the transformation*) to extract data from source systems (*external computer systems*) and store it as an event log (*process protocol*).** Rinke937 teaches its computer system executing a process sensor to extract data from “raw data stored in the source systems” (*data extracted from the data tables stored in the external computer systems*) and store that extracted data in an event log (*process protocol*). EX1004, [0090]. For example, Rinke937 describes the computer system processor “executing ... at least one process sensor” to “derive process data from the [source system’s] raw data,” generate data from that derived raw data, and store the generated data as an “event log.” *Id.*, [0012]-[0020]; *see also id.*, [0021]-[0023], [0010]. Using the Rinke937 process sensors, raw data from a source system “is transformed into a representation which is capable for process mining applications.” *Id.*, [0072]; *see also supra* Section IX.E.7 (explaining creation of the process sensors). Rinke937 also expressly contemplates reading raw data from source systems “in a scheduled manner,” including daily execution of process sensors. EX1004, [0113].

140. Rinke937 also teaches multiple “source systems” storing raw data (*external computer systems*). EX1004, [0090] (discussing “raw data stored in the source systems”), [0103] (discussing “source systems”). Rinke937 teaches that a process sensor obtains process-step information from “raw data stored in the source systems,” using table definitions and joins to derive the object identifier, process-step description, and ordering information. EX1004, [0090], [0096]-[0102]. Harrison also teaches extracting data from tables on multiple source database systems, and a network interface connecting the processing system to those sources. EX1006, [0055], [0456]; *supra* Section IX.E.1.a. Harrison expressly describes “connecting to the source database” and “read[ing] the contents of the table from the source database.” EX1006, [0081], [0087]-[0089].

141. In the Rinke937-Nooijen-Harrison combination, the computer system executes process sensors—created as described above for element 1.7—to extract data from the data tables stored in at least two source systems (*external computer systems ES1, ES2*). *Supra* Section IX.D. Rinke937-Nooijen-Harrison combines the source connections and related teachings of Harrison with process sensor execution, as taught by Rinke937, to retrieve data from source systems and store it in event logs for use by a process mining system. Rinke937 expressly contemplates reading raw data stored in source systems “in a scheduled manner,” including daily execution. EX1004, [0113]. Once the mappings identify the relevant tables, fields, and

relationships, subsequent executions use those definitions to retrieve object identifiers and associated event data without repeating the broader search for relationships among candidate tables and columns. EX1005, 319-324; EX1004, [0096]-[0102], [0117]-[0119]. Reading the original source tables also captures newly recorded activities without requiring a separate refresh of the entire locally imported table collection. Using the imported copies for the computationally intensive discovery and process sensor creation, then directing subsequent process sensor executions to the original source tables, permits the configured process sensors to capture newly recorded activities without making event-log generation depend on a separate, and potentially time-consuming, refresh of the local imported data tables. *Supra* Section IX.D.

142. Harrison preserves the information needed to direct those reads to the original tables. Its import metadata identifies the source system or database, table name, original column name, and any renamed column in the imported copy. EX1006, [0111]-[0117]. A POSITA would have used that metadata to translate locally determined table and field references in Nooijen into the corresponding original source references when generating the executable sensor statements taught by Rinke⁹³⁷ so that the executable process sensors access and read raw data from the target tables in the source systems. EX1004, [0117]-[0119]; EX1005, 321-24.

The metadata thus identifies where to retrieve the mapped fields, while the reconstructed relationships specify how to associate their values.

143. During execution, the local computer system would use the source connections taught by Harrison to retrieve the required identifier, activity, and timestamp fields from the corresponding external tables. EX1006, [0087]-[0090], [0456]. Where related tables reside in different external systems, the local Rinke937-Nooijen-Harrison system would retrieve the required fields separately and join the retrieved records using the key-column correspondences specified in the mappings. For example, matching an object identifier retrieved from the main table with the corresponding foreign-key value retrieved from an event table associates that event with the identified object; paths through intermediate tables provide additional matching columns where needed. These are the key-based joining and event-association operations taught by Nooijen and used to define the executable process sensor units in Rinke937. EX1005, 323-324; EX1004, [0099]-[0102]. In this implementation, the local Rinke937-Nooijen-Harrison system executes the transformation, retrieves the source data, and combines the retrieved records according to the locally determined mappings.

144. **Rinke937-Nooijen-Harrison stores the extracted data as the process protocol.** Rinke937 expressly teaches “storing, by the processor, the generated data as event log with the storage means.” EX1004, [0020]. Its event-log

structure includes the process-object identifier, process-step identifier, and ordering information, including timestamps, generated by the process sensors. *Id.*, [0021]-[0026]; *see also id.*, [0075]-[0085], [0097]-[0102]. Rinke937-Nooijen-Harrison retains those operations to store the activity data obtained from both external systems as the common event log (*process protocol*), associating each recorded process step with its object and order. *Supra* Sections IX.D, IX.E.1.a.

145. **Rinke937-Nooijen-Harrison provides the created process protocol to a Process Mining System for analysis and reconstruction of the processes carried out in the external computer systems.** Rinke937 expressly states that “[t]he generated data stored with the storage means may be provided to a process mining system,” and that, “[g]iven such an event log, the as-is process may be reconstructed by process mining algorithms.” EX1004, [0035], [0088]; *see also id.*, [0001], [0008], [0010], [0039], [0071]-[0080], [0120]-[0122], [0137]-[0140]. In Rinke937-Nooijen-Harrison, the event log contains the object-related process steps extracted from the original tables in the first and second external computer systems. Providing the log therefore supplies the Process Mining System with the recorded activities used to analyze and reconstruct the processes carried out in those systems. *Supra* Sections IX.D, IX.E.1.a, IX.E.1.d. Accordingly, Rinke937-Nooijen-Harrison teaches element [1.8].

146. Thus, Rinke937-Nooijen-Harrison teaches each element of claim 1 and their claimed combination and renders claim 1 obvious.

F. Claim 2

[2] The method of claim 1, wherein at least one of the selected data tables is imported from the first external computer system (ES1), and at least one of the selected data tables is imported from the second external computer system (ES2).

147. Rinke937-Nooijen-Harrison renders claim 2 obvious. As explained for claim 1, Rinke937-Nooijen-Harrison imports tables from first and second external computer systems as taught by Harrison, then reconstructs their relationships, and selects a main table (*first data table*) and related event tables (*second data tables*) for creating the event log, as taught by Nooijen. *Supra* Sections IX.D, IX.E.2-IX.E.3. POSITAs would understand that applying these techniques to related process data distributed across two external computer systems would include selecting at least one table imported from each external system, as explained below.

148. **Rinke937-Nooijen-Harrison selects related tables imported from different external computer systems.** Harrison teaches importing tables “from any number of data sources” and explains that “it may often be the case that data imported from disparate data sources is interrelated.” EX1006, [0055], [0058]-[0059], [0065], Fig. 1; *see also id.*, [0003] (contemplating “relationships between different data sets” from “diverse data sources”), [0005]-[0006]. Rinke937-Nooijen-Harrison applies the relationship-discovery techniques taught by Harrison and

Nooijen to those imported tables, then uses the reconstructed relationships to select the main table and related event tables. *Supra* Sections IX.D, IX.E.3-IX.E.5. Nooijen selects event tables based on timestamps and foreign-key relationships and identifies the paths connecting those tables to the main table. EX1005, 321-323. Where relevant event information resides in tables imported from different external systems, these relationships allow the Rinke937-Nooijen-Harrison system to select tables from both systems for the same event log. POSITAs would recognize that in at least some obvious use cases, the main table (*first data table*) and one or more of the event tables (*second data tables*) would come from “disparate” but “interrelated” data sources. EX1006, [0055].

149. Harrison provides a concrete example of related data distributed across different sources. A “provider may import a database of gas supply accounts from a supply part of the organisation and a database of boiler maintenance data from a service / maintenance part of the organisation.” EX1006, [0065]. Harrison explains that “[t]he data may be related in that some supply customers may also be maintenance customers.” *Id.* Applying Rinke937-Nooijen-Harrison to analyze activities for those shared customers would have made it obvious to select a customer-identifying main table from the gas-supply system (*at least one selected data table imported from the first external computer system*) and a related event table recording maintenance activities from the service/maintenance system (*at least one*

selected data table imported from the second external computer system). The reconstructed relationships would associate the maintenance events with the corresponding customers, consistent with Nooijen's teaching of joining event information with the main table to associate each event with a trace. EX1005, 324.

150. A POSITA would have been motivated to select these relevant tables from both systems to include the relevant activities for the shared customers in the resulting event log. Restricting selection to tables from only one system would omit relevant activity information maintained in the other system. Selecting tables from both systems would predictably apply the table selection and relationship-mapping techniques taught by Nooijen. Accordingly, Rinke937-Nooijen-Harrison teaches the additional limitation of claim 2 and renders this claim obvious.

G. Claim 3

[3] The method of claim 1, wherein during reconstructing of the relationships between the imported data tables in the step (b), it is determined via which attributes the data tables are related to each other.

151. Rinke937-Nooijen-Harrison teaches and renders obvious this claim. As explained above, Rinke937-Nooijen-Harrison teaches each limitation of claim 1, including determining and reconstructing relationships between the locally imported tables, and thus renders claim 1 obvious. *Supra* Section IX.E.

152. Further, during determination and reconstruction of the relationships between the imported data tables (*supra* Section IX.E), the combined Rinke937-

Nooijen-Harrison system determines the attributes by which the imported data tables are related. Nooijen teaches reconstructing a relational schema containing the table schemas and foreign-key relationships. EX1005, 319. Each foreign key is defined by the parent table and its primary-key columns (*attributes*), together with the child table and its referencing columns (*attributes*). *Id.* Identifying a foreign-key relationship therefore identifies both the related tables and the particular columns through which their records are linked. Nooijen expressly applies this discovery to “undocumented foreign key relations” (*id.*) and thus determines the attributes by which the tables are related. This complements Harrison, which teaches to identify “relationships between table columns” based on whether the columns contain keys and the overlap between their data contents. EX1006, [0068]. Harrison’s “Table Analyser” examines imported table contents to identify a relationship between a primary key in one table and a corresponding foreign key in another. *Id.*, [0221]. As in Nooijen, the selected tables are related by the identified columns (*attributes*).

153. The identified columns and foreign-key relations are thus determined attributes through which the imported tables are related. *See* EX1004, [0096]-[0102]; *supra* Section IX.D. Accordingly, Rinke937-Nooijen-Harrison renders this claim obvious.

H. Claim 4

[4] The method of claim 3, wherein two data tables are related to each other, if an attribute of the one data table is identical or similar to an attribute of the other data table.

154. Rinke937-Nooijen-Harrison teaches that two data tables are related if an attribute of one data table is identical or similar to an attribute of the other data table, and thus renders this claim obvious. As explained for claim 3, Rinke937-Nooijen-Harrison reconstructs relationships between the imported data tables by identifying the columns (*attributes*) through which those tables are related. *Supra* Section IX.G. Rinke937-Nooijen-Harrison incorporates the techniques in Harrison for discovering relationships between imported tables and the techniques in Nooijen for representing those relationships as primary and foreign key relations, which support the table joins used to generate the process sensors in Rinke937. *Supra* Sections IX.D, IX.E.3.

155. Rinke937-Nooijen-Harrison identifies a relationship between two tables having identical customer identifier attributes. Harrison provides a concrete example, where tables 2302 and 2304 “may have originated from different data sources,” and the Table Analyser discovers relationships between their columns. EX1006, [0421], Fig. 23A. Harrison specifically identifies “a relationship between the CUSTID column of table 2302 and the CUSTID column of table 2304” as the strongest proposed relationship. EX1006, [0421]. Thus, the CUSTID column of

table 2302 (*attribute of one data table*) and the CUSTID column of table 2304 (*identical attribute of the other data table*) provide the relationship between the two tables. This example satisfies the claimed “identical or similar” alternative through identical customer identifier attributes.

156. Further, Harrison teaches that the Table Analyser considers the probability that columns are keys and the degree of overlap between their values to identify relationships, including primary and foreign key relationships. EX1006, [0068], [0231]-[0233]. Harrison separately illustrates a customer identifier column in a customers table functioning as a primary key and a customer identifier column in an orders table functioning as a foreign key. EX1006, [0232]. These disclosures establish a relationship through shared customer identifier attributes and corresponding key data, rather than relying solely on matching column labels.

157. Rinke937-Nooijen-Harrison uses the identified attributes to reconstruct and apply the relationship between the tables. Nooijen represents a foreign key relation using the primary key columns of one table and the corresponding referencing columns of another table (*attributes through which the tables are related*). EX1005, 319. Harrison likewise uses the discovered column relationships as the basis for table joins, and Rinke937 uses joins between related tables when defining the transformations incorporated into process sensors. EX1006, [0235], [0422]; EX1004, [0099]-[0102]. As explained in the combination section and for

claim 3, these teachings operate together: Harrison identifies the relationship through the customer identifier attributes, Nooijen represents that relationship as a key relation, and Rinke937 uses the related columns in its table joins. *Supra* Sections IX.D, IX.G. Accordingly, Rinke937-Nooijen-Harrison teaches and renders claim 4 obvious.

I. Claim 5

[5] The method of claim 1, wherein the relationships reconstructed between two imported data tables are stored, wherein the names of the two data tables, the respective attributes, and an assignment of the attributes of the one data table to the attributes of the other data table are stored.

158. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches each element of claim 1, including determining and reconstructing relationships between the locally imported tables, and thus renders claim 1 obvious. *Supra* Section IX.E, IX.E.3. Nooijen represents the reconstructed relationships in a relational schema containing the table schemas and foreign keys. EX1005, 319. Harrison expressly teaches generating “an output file 906” containing “information on identified table relationships.” EX1006, [0234]. Its detailed example stores entries identifying the tables, their columns, and the column pairs through which the tables may be related. *Id.*, [0289]-[0296]. These teachings provide a stored representation of the reconstructed relationships for subsequent mapping and extraction.

159. Nooijen represents each reconstructed foreign key as $F=(T_p, C_p, T_c, C_c)$, identifying the parent table T_p and child table T_c (*the two related imported data tables*), the primary key columns C_p and referencing columns C_c (*the respective attributes*), and the foreign key relation linking those columns (*the assignment between the attributes*). EX1005, 319. The reconstructed schema therefore identifies both sides of each relationship and which attributes link them. Harrison reinforces this representation by storing identifiers for the tables and columns together with entries for related column pairs. EX1006, [0290]-[0293].

160. A POSITA would have implemented this stored representation using the names of the related tables and columns so that subsequent queries could identify and join the correct tables through the correct attributes. Harrison explains that its illustrative column labels represent “named columns” and permits “any appropriate encoding or referencing” to identify the source table and column. EX1006, [0240], [0245]. Nooijen uses the reconstructed schema to generate mappings containing the main table and links to other tables “in the form of foreign keys,” and those mappings are used to construct database queries. EX1005, 321. Recording the table names, column names, and their pairings would thus have been a predictable way to preserve the discovered relationships in a form usable by the subsequent extraction operations in Rinke937-Nooijen-Harrison.

161. Rinke937-Nooijen-Harrison retains this relationship information for generating and executing the process sensors. As explained in the combination section, the reconstructed schema and automated mappings in Nooijen provide the table relationships used within the event log generation framework of Rinke937. *Supra* Section IX.D. Rinke937 expressly scripts the “relations, source tables, target tables and other attributes” in the sensor definitions and generates executable code from those definitions, and its event log package contains a data model describing the structure of the elements and additional tables linked to the event log. EX1004, [0099], [0117]-[0119]. The stored table names, attributes, and attribute assignments therefore support the joins performed by the process sensors when extracting the process data. *Id.*, [0099]-[0102]. A POSITA therefore would have understood the reconstructed relationship information taught by Nooijen to be stored as part of the schema or data model used to perform the subsequent event log extraction. Accordingly, Rinke937-Nooijen-Harrison teaches and renders claim 5 obvious.

J. Claim 6

[6] The method of claim 1, wherein a first data model belonging to the data tables imported from the first external computer system (EC1) is imported from the first external computer system into the local computer system, wherein the relationships between the data tables imported from the first external computer system (EC1) are reconstructed based on the first data model, a second data model belonging to the data tables imported from the second external computer system (EC2) is imported from the second external computer system into the local computer system, wherein the relationships between the data tables imported from the second external computer system (EC2) are reconstructed

based on the second data model, such that only the relationships between the data tables from the first external computer system and the data tables from the second external computer system have to be reconstructed without the use of a data model.

162. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. As explained above, Rinke937-Nooijen-Harrison teaches each element of claim 1, including importing tables from multiple external computer systems and reconstructing their relationships locally, and thus renders claim 1 obvious. *Supra* Sections IX.D, IX.E, IX.E.2-IX.E.3.

163. Further, **Rinke937-Nooijen-Harrison imports a data model from each external computer system into the local computer system.** Harrison teaches connecting to a source database and extracting its table metadata into metadata files. EX1006, [0087], Fig. 2B. Its metadata generator reads schema information from different source databases and stores that information in a metadata repository. *Id.*, [0103], [0111]-[0125]. Harrison expressly contemplates using known relationships where “relationship information was extracted from the source database.” *Id.*, [0382]. Applying these teachings to the two source systems imports the schema metadata and relationship information from the first source database (*first data model*) and the corresponding information from the second source database (*second data model*) into the Rinke937-Nooijen-Harrison computer system (*local computer system*).

164. **Rinke937-Nooijen-Harrison uses each imported data model to reconstruct relationships among the tables imported from the corresponding external system.** Nooijen explains that schema information identifies table columns, primary keys, and dependencies documented as foreign key relations. EX1005, 319. Its relational schema represents each foreign key through the parent table and its primary key columns, together with the child table and its referencing columns. *Id.* Those definitions specify which tables are related and the attributes through which they are linked. Harrison accommodates both relationships extracted from the source database and relationships discovered through analysis of imported table contents. EX1006, [0382]-[0383]. In Rinke937-Nooijen-Harrison, the imported schema and relationship definitions from each source system are used to establish a local representation of the relationships among its imported tables: the first source schema (*first data model*) provides the relationships among tables from the first external system, and the second source schema (*second data model*) provides the relationships among tables from the second external system.

165. **Rinke937-Nooijen-Harrison uses the imported models for internal relationships so that only relationships between tables from different external systems require reconstruction without a data model.** Nooijen identifies incomplete schema information as the reason additional relationship discovery “may be required,” and specifically discovers “undocumented foreign key relations.”

EX1005, 319. A POSITA would have applied that discovery to relationships missing from the imported models, while using the available model definitions for relationships already documented. For source systems whose models document their internal table relationships, this leaves relationships across the two systems to be reconstructed without a source model. Like Nooijen, Harrison also teaches discovering relationships from disparate sources, as Harrison analyzes imported table contents to identify relationships where related tables “originated from different data sources.” EX1006, [0068], [0221], [0421].

166. This implementation of Rinke937-Nooijen-Harrison avoids unnecessary rediscovery of documented relationships and concentrates the analysis on the relationships absent from the source models. Harrison similarly teaches limiting analysis to new relationships to avoid fully analyzing all possible column combinations. EX1006, [0314]. The resulting combined schema provides the relationships used by the automated mappings in Nooijen and the process sensors in Rinke937, as explained above. *Supra* Sections IX.D, IX.E.3, IX.E.7. Accordingly, Rinke937-Nooijen-Harrison renders claim 6 obvious.

K. Claim 7

[7] The method of claim 1, wherein during determining and selecting of a first data table in the step (c), it is determined, which data table meets one or more conditions, wherein the conditions are selected from a group, consisting of the data table comprises less data sets compared to the other data tables, the data

table comprises an attribute, in which the time stamps are stored, the data table is referenced by a few other data tables, and combinations thereof.

167. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 1 obvious. *Supra* Section IX.E. Further, as explained for claim 1, Rinke937-Nooijen-Harrison imports the source tables, reconstructs their relationships, and applies the automated selection in Nooijen to identify a main table (*selected first data table*) whose records identify the objects for which the event log (*process protocol*) will be created. *Supra* Sections IX.D, IX.E.2-IX.E.4. Claim 7 permits selection of the first data table based on one of the listed conditions. Rinke937-Nooijen-Harrison uses foreign-key relationships to select a main table, as taught by Nooijen, thus teaching the condition that *the data table is referenced by a few other data tables* as discussed below.

168. **Rinke937-Nooijen-Harrison uses foreign key references from related auxiliary tables to identify the main table.** Nooijen explains that business objects materialize as important tables, while their additional attributes materialize in related auxiliary tables. EX1005, 320. Its selection technique groups these tables into artifact schemas using a “distance function between tables” that considers table “importance and foreign key relations.” *Id.* Specifically, “[f]oreign key relations between tables associate auxiliary tables to important tables.” *Id.* Nooijen defines those relations through a parent table and its primary key columns, together with a

child table and its referencing columns. *Id.*, 319. Thus, auxiliary child tables containing foreign keys reference the important parent table through its primary key. Those foreign key references contribute to the distances used to cluster the tables and select “a unique main table” having “the least distance to all other tables in the cluster.” *Id.*, 320. The foreign key references therefore are used in determining and selecting the main table.

169. **Rinke937-Nooijen-Harrison selects a main table that is referenced by a few related auxiliary tables.** The technique in Nooijen selects the main table from a group of related tables. EX1005, 320. The related auxiliary tables reference the main table through the foreign-key relationships considered in the selection analysis (*the data table is referenced by a few other data tables*). *Id.* Nooijen provides a concrete example of applying this selection technique to a small group of related tables. Its evaluated artifact schema A2 contains six tables, including the main table. EX1005, 325. Nooijen explains that the tables constituting an artifact form a cluster, that foreign-key relationships associate auxiliary tables with important tables, and that the main table is selected using distances reflecting those relationships and table importance. *Id.*, 320. The A2 schema therefore illustrates application of that relationship-based selection to a main table and five other tables within the artifact schema. Applying the selection taught by Nooijen to this small group identifies the main table (*selected first data table*) based in part on its foreign-

key relationships with the relevant auxiliary tables (*other data tables referencing the first data table*). A POSITA would have understood this implementation to teach the condition that the selected table *is referenced by a few other data tables*.

170. Rinke937-Nooijen-Harrison uses the locally imported tables for this analysis, as taught by Harrison, to select the main table using the reconstructed relationships, as taught by Nooijen, and uses the resulting table and relationship definitions to generate its process sensor units, as taught by Rinke937. *Supra* Sections IX.D, IX.E.3-IX.E.4, IX.E.7. Accordingly, Rinke937-Nooijen-Harrison renders claim 7 obvious.

L. Claim 8

[8] The method of claim 7, wherein information with respect to the selected first data table is stored.

171. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 7 obvious. *Supra* Section IX.K.

172. Further, Rinke937-Nooijen-Harrison stores information with respect to the main table (*selected first data table*). Nooijen teaches recording information regarding the main table in the generated schema-log mapping, including the main table, its trace identifier columns, foreign key relations to other tables, and associated attribute mappings (*information with respect to the selected first table*). EX1005,

320-22. Further, the trace mapping, which includes the specification of the main table, is stored for subsequent query generation and event extraction. *Id.*, 324. Rinke937 further teaches “storing auxiliary data” and maintaining a data model describing the tables and relationships used to create the event log. EX1004, [0036]-[0039]. In Rinke937-Nooijen-Harrison, the information taught by Nooijen is stored, as taught by Rinke937, to provide the definitions for the source table and relationships to the process sensor framework taught by Rinke937. *Id.*, [0036]-[0039], [0117]-[0119]; *supra* Section IX.D. Accordingly, Rinke937-Nooijen-Harrison renders claim 8 obvious.

M. Claim 9

[9] The method of claim 1, wherein during determining and selecting of the number of second data tables in the step (d) those data tables are selected, which meet one or more conditions, wherein the conditions are selected from the group, consisting of the data tables comprise many data sets compared to the first data table, the data table is referenced directly or indirectly by the first data table, the data tables respectively comprise an attribute, in which the time stamps are stored, and combinations thereof.

173. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 1 obvious. *Supra* Section IX.E. As explained for claim 1, Rinke937-Nooijen-Harrison applies the automated mapping techniques taught by Nooijen to select related event tables (*second data tables*) from the locally imported tables based on the main table (*first data table*) and the reconstructed relationships. *Supra* Sections IX.D, IX.E.5.

174. **Rinke937-Nooijen-Harrison selects second data tables that are referenced directly or indirectly by the first data table, as taught by Nooijen.** Nooijen uses foreign key relationships to associate auxiliary tables with the main table and identify the tables belonging to the artifact schema. EX1005, 319-320. In determining the relationships to the main table, Nooijen expressly describes “recursively verifying foreign keys in the child direction and the parent direction.” *Id.*, 323. In the child direction, the algorithm checks whether multiple records in the child table correspond to each record in the base table. *Id.* Thus, Nooijen uses foreign-key relationships to identify related records in either direction, including starting from a record in the main table and identifying corresponding records in a child table.

175. A POSITA would have understood why this works. Where an event table contains a foreign key identifying an object in the main table, matching the foreign-key values to the corresponding main-table key values permits the system both to identify the object associated with a particular event and to identify the events associated with a particular object. The relationship supplies the matching columns needed to identify and join records in either direction. Starting from the main table, the system can therefore identify the related event records even when the foreign-key attribute resides in the event table.

176. Nooijen further teaches using these relationships through both direct and indirect connections. For each selected event table T_{ET} (*second data table*), the mapping algorithm determines foreign keys $F_{link} = \text{Path}(T_m, T_{ET})$, linking the main table T_M (*first data table*) to that event table. *Id.*, 321-322. Nooijen explains that this “path between two tables [] consists of the sequence of foreign keys connecting those tables.” *Id.*, 323. A single foreign-key relationship *directly* connects the main table to the event table; a sequence of foreign-key relationships connects them *indirectly* through intervening tables. Nooijen uses these paths to join the event table with the main table, including intervening tables where needed, and associate the extracted events with the corresponding object. *Id.*, 324.

177. A POSITA would have understood that these connections satisfy the condition that the second data table is “referenced directly or indirectly by the first data table,” consistent with the use of relationships in the ’708 patent. The ’708 patent identifies “Order Number” in table T1 as a primary-key attribute and “Reference ID” in table T3 as a foreign-key attribute, establishing the relationship “T1 (Order Number) $\leftrightarrow$ T3 (Reference ID).” EX1001, 13:6-14. The ’708 patent also explains that “Tn (x) $\leftrightarrow$ Tm (y) means that the tables Tn and Tm are linked to each other via the attribute x in the table Tn and via the attribute y in the table Tm.” EX1001, 12:20-22. The ’708 patent selects T1 as the first data table and selects T3 as a second data table because it is directly related to T1 and stores actions relevant

to the process protocol. *Id.*, 14:1-12. This example confirms that the “referenced directly or indirectly” condition encompasses identifying a related second data table from the first data table even where the foreign-key attribute resides in the second data table. Nooijen teaches that operation independently: its foreign-key paths identify event tables connected to the main table, directly or indirectly, and thus enable their event records to be associated with the objects for which the event log is created. Thus, Rinke937-Nooijen-Harrison selects event tables linked to the main table through foreign key paths (*data tables referenced directly or indirectly by the first data table*).

178. **Rinke937-Nooijen-Harrison also selects second data tables based on those tables containing attributes in which timestamps are stored, as taught by Nooijen.** Nooijen expressly identifies event types by “selecting all columns with a timestamp domain as event columns” and defines the corresponding event table (T_{ET}) as the table containing that timestamp column. EX1005, 322-323. For each selected event column C_e (*attribute in which timestamps are stored*), Nooijen identifies the event table T_{ET} (*selected second data table*) containing that column. *Id.*, 322-23. Thus, the timestamp attribute is an additional criterion used to identify and select the event table, teaching the condition *the data tables respectively comprise an attribute, in which the time stamps are stored*.

179. Rinke937-Nooijen-Harrison therefore teaches the additional element recited in claim 9 and renders claim 9 obvious.

N. Claim 10

[10] The method of claim 9, wherein information with respect to the selected second data tables and with respect to the relationships between the selected second data tables and the first data table are stored.

180. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 9 obvious. *Supra* Section IX.M.

181. Further, Rinke937-Nooijen-Harrison stores information regarding the selected second data tables and their relationships with the selected first data table. Nooijen records the event identifier columns and timestamp column (*information regarding each selected second data table*), the main table (*selected first data table*), and the foreign key relations connecting those tables (*relationships between the selected second data tables and the first data table*) in its generated event mappings. EX1005, 321-322. As explained for claim 9, these paths connect each selected event table to the main table either directly or through intervening tables. *Id.*, 322-323; *supra* Section IX.M. The mappings are retained and supplied to the extraction tool to generate the database queries that join the identified tables using the recorded relationships. EX1005, 324. In Rinke937-Nooijen-Harrison, this stored information provides the table and relationship definitions used to generate the executable

process sensor units in Rinke937, as explained for element [1.7]. EX1004, [0102]; *supra* Sections IX.D, IX.E.7. Accordingly, Rinke937-Nooijen-Harrison teaches the additional element recited in claim 10 and renders claim 10 obvious.

O. Claim 11

[11] The method of claim 1, wherein during creating of the executable transformation in the step (f) it is determined, which attributes of the selected first data table and of the selected second data tables are considered during extracting of the data.

182. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 1 obvious. *Supra* Section IX.E.

183. Further, Rinke937-Nooijen-Harrison determines which attributes of the selected first and second data tables will be considered during extraction when creating the executable transformation. Nooijen generates mappings identifying the trace identifier columns of the main table (*attributes of the selected first data table*) and the event identifier and timestamp columns of the event tables (*attributes of the selected second data tables*). EX1005, 321-22. It also assigns additional columns as instance or event attributes and creates attribute mappings specifying each attribute name, type, and source column C_a (*attribute considered during extraction*). *Id.*, 321-323. These mappings therefore identify the particular columns to use before the extraction queries are generated. *Id.*, 324.

184. In Rinke937-Nooijen-Harrison, those column determinations supply the field definitions used to create the process sensor code (*executable transformation*), as explained for element [1.7]. *Supra* Sections IX.D, IX.E.7. Rinke937 expressly defines the source-table fields supplying the object identifier, process-step information, timestamp, and additional event-log attributes, and generates executable code from the scripted or graphical sensor definitions. EX1004, [0099]-[0102]; *id.*, [0117]-[0119]. The attributes identified through the mappings in Nooijen are incorporated into the transformation during its creation, thus teaching the additional element of claim 11 and rendering this claim obvious.

P. Claim 12

[12] The method of claim 11, wherein during executing of the transformation in the step (g) data sets are read out from the first data table and from the second data tables, wherein the data sets, which have been read out, comprise the attribute values of the attributes determined for the respective data table, wherein the data sets, which have been read out, are stored as process protocol, wherein each data set stored in the process protocol represents a process step (Event) of a process executed in the first external computer system and in the second external computer system.

185. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 11 obvious. *Supra* Section IX.O.

186. Further, **Rinke937-Nooijen-Harrison reads data sets containing the determined attribute values from the selected first and second data tables**

during execution of the process sensor code (*transformation*). As explained for claim 11, the automated mappings in Nooijen provide the table and attribute definitions incorporated into the executable process sensor code. *Supra* Section IX.O. Rinke937 teaches that the sensor definition “is evaluated during the execution of the sensor,” retrieving values from the defined fields through joins across the source tables. EX1004, [0096]-[0102]. The process sensors thus read the object identifier fields from the main table (*selected first data table*) and the process step, timestamp, and additional attribute fields from the related tables (*selected second data tables*), as specified during creation of the executable process sensor code. In Rinke937-Nooijen-Harrison, the process sensor code reads the raw data in the external source systems, as explained for element [1.8]. *Supra* Section IX.E.8.

187. **Rinke937-Nooijen-Harrison further stores the resulting data sets read out from the first and second data tables as an event log (*process protocol*), with each stored record representing a process step.** Rinke937 expressly teaches storing the data generated during process sensor execution as an event log. EX1004, [0014]-[0023]. Rinke937 further specifies that the identifier of a single process step, the associated object identifier, and the ordering information form “a single data record” (*data set representing a process step*). *Id.*, [0025]. The executable process sensors of Rinke937-Nooijen-Harrison contribute records reflecting the process steps carried out across the first and second external systems to the complete event

log, as explained for claim 1. *Supra* Sections IX.E.1, IX.E.8. Accordingly, Rinke937-Nooijen-Harrison teaches the additional element of claim 12 and renders this claim obvious.

Q. Claim 13

[13] The method of claim 12, wherein the data sets, which have been read out, are stored according to a predetermined data structure in the process protocol, wherein the predetermined data structure comprises at least: a first attribute, in which a first unique identifier of the process of the respective process step is stored, a second attribute, in which an identifier of the respective process step is stored; and a third attribute, in which a sequence of the process steps within a process is stored.

188. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 12 obvious. *Supra* Section IX.P.

189. Further, as discussed above, Rinke937 discloses the same data structure recited in claim 13, as Rinke expressly discloses that the event log (*process protocol*) has a “predetermined data structure comprising at least a first attribute for storing the unique identifier of the process element of the respective process step, a second attribute for storing the identifier of the respective process step, and a third attribute for storing the order of the process steps within a process element.” EX1004, [0020]-[0023]; *id.*, [0025]-[0026], [0081]-[0086]; *supra* Section IX.E.1.a. Rinke937-Nooijen-Harrison uses this “predetermined data structure” taught by Rinke937 for the event log when storing the records generated by its process sensors, as explained

for claim 12. *Supra* Section IX.P. Accordingly, Rinke937-Nooijen-Harrison teaches the additional element of claim 13 and renders this claim obvious.

R. Claim 14

[14] The method of claim 13, wherein the attribute values of the attributes of the process protocol are determined from the attribute values of the data sets, which have been read out, and/or the table names and/or column names of the selected first data table and/or the selected second data tables.

190. Rinke937-Nooijen-Harrison teaches and renders this claim obvious. First, Rinke937-Nooijen-Harrison teaches and renders claim 13 obvious. *Supra* Section IX.Q.

191. Further, **Rinke937-Nooijen-Harrison determines the attribute values of the attributes of the event log from the attribute values read from the selected data tables.** Rinke937 teaches deriving the object identifier, process step information, timestamp, and “further attributes” for the event log (*attribute values of the process protocol*) from the fields of the source table, including through transformations and joins across related tables. EX1004, [0099]-[0102]. As a concrete example, the “Change of content sensor” taught by Rinke937 retrieves the “unique identifier” from the invoice table BSEG (*selected first data table*), the “process step” from the change log table CDPOS, and the “timestamp” from the change log table CDHDR (*selected second data tables*), and stores the resulting values in the event log (*process protocol*). *Id.*, [0130]-[0136].

192. In Rinke937-Nooijen-Harrison, the mappings in Nooijen identify the fields used by the process sensors, as explained for claims 11-13. *Supra* Sections IX.O-IX.Q. Accordingly, Rinke937-Nooijen-Harrison teaches the “attribute values” alternative recited in claim 14, and renders claim 14 obvious.

I, Craig A. Knoblock, do hereby declare and state, that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: October 21, 2026


Signature

APPENDIX A

Craig A. Knoblock

University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
Email: knoblock@isi.edu
Phone: (310) 448-8786
<http://www.isi.edu/~knoblock>
October 5, 2026

Education

Carnegie Mellon University, Pittsburgh, PA, 1984-1991
Ph.D., Computer Science, May 1991

Carnegie Mellon University, Pittsburgh, PA, 1984-1988
M.S., Computer Science, May 1988

Syracuse University, Syracuse, NY, 1980-1984
B.S., Computer Science with Honors, May 1984

Research & Administrative Positions

Distinguished Principal Scientist, July 2026 – present
Keston Executive Director, Information Sciences Institute, April 2020 – June 2026
Interim Keston Executive Director, Information Sciences Institute, June 2018 – April 2020
Division Director, Intelligent Systems Division, October 2017 – June 2018
Research Director, Center on Knowledge Graphs, June 2010 – June 2018
Senior Project Leader, May 2001 – June 2010
Project Leader, May 1997 – May 2001
Senior Research Scientist, August 1991 – May 1997
Information Sciences Institute
University of Southern California
Research Professor, April 2006 – present
Research Associate Professor, November 1998 – March 2006
Research Assistant Professor, September 1991 – November 1998
Computer Science Department
University of Southern California
Research Professor, January 2014 – present
Affiliated Faculty Member, 2010 – January 2014
Spatial Sciences Institute
University of Southern California
Vice Dean of Engineering, April 2020 – June 2026
Viterbi School of Engineering
University of Southern California
Director of Data Science Program, November 2018 – April 2020
Associate Director of Data Informatics, October 2014 – November 2018
University of Southern California

Key Investigator, September 1996 – September 2007

Integrated Media Systems Center

University of Southern California

Cofounder & Chief Scientist, November 1999 – December 2010

Fetch Technologies, Inc.

Acquired by Connotate, Inc. in December, 2011

Strategic Advisor, June 2012 – July 2014

Cofounder & Chief Scientist, November 2004 – June 2012

Geosemble Technologies, Inc.

Acquired by TerraGo Technologies in June 2012

Research Interests

Research focus: Techniques for describing, acquiring, and exploiting the semantics of data: source modeling, schema alignment, ontology alignment, entity linking, data cleaning, information extraction, web data extraction, and combining all of these techniques to create linked data and build knowledge graphs.

Teaching Experience

- **DSCI 558: Building Knowledge Graphs**
Taught jointly with Professor Jay Pujara, University of Southern California, Spring 2021
- **Short course on Building Knowledge Graphs**
BigDat 2020: 6th International Winter School on Big Data, Ancona, Italy, Jan 2020
- **Short course on Building Knowledge Graphs**
BigDat 2019: 5th International Winter School on Big Data, Cambridge, United Kingdom, Jan 2019
- **INF 558: Building Knowledge Graphs**
Taught jointly with Professor Pedro Szekely, University of Southern California, Fall 2017
- **CSCI 548: Information Integration on the Web**
Two sections taught jointly with Professor Pedro Szekely, University of Southern California, Fall 2015
- **CSCI 648: Advanced Information Integration**
Taught jointly with Professor Jose Luis Ambite, University of Southern California, Spring 2015
- **CSCI 548: Information Integration on the Web**
Taught jointly with Professor Jose Luis Ambite, University of Southern California, Spring 2015
- **CSCI 548: Information Integration on the Web**
Two sections taught jointly with Professor Pedro Szekely, University of Southern California, Fall 2014
- **CSCI 599: Geospatial Data Integration**
Taught jointly with Professor Yao-Yi Chiang, University of Southern California, Spring 2014.
- **CSCI 548: Information Integration on the Web**
Taught jointly with Professor Pedro Szekely, University of Southern California, Fall 2013.
- **CSCI 548: Information Integration on the Web**
Taught jointly with Professors Pedro Szekely and Jose Luis Ambite, University of Southern California, Spring 2012.
- **CSCI 599: Geospatial Data Integration**
University of Southern California, Spring 2011.
- **Geospatial Data Integration**
University of Trento, Italy, July 2010.

- **CSCI 548: Information Integration on the Web**
Taught jointly with Professor Kristina Lerman, University of Southern California, Spring 2010.
- **CSCI 548: Information Integration on the Web**
Taught jointly with Professor Kristina Lerman, University of Southern California, Spring 2009.
- **CSCI 548: Information Integration on the Web**
University of Southern California, Spring 2008.
- **CSCI 548: Information Integration on the Web**
University of Southern California, Fall 2006.
- **CSCI 548: Information Integration on the Web**
University of Southern California, Spring 2005.
- **CSCI 548: Information Integration on the Web**
University of Southern California, Spring 2004.
- **CSCI 599: Information Integration on the Web**
University of Southern California, Spring 2003.
- **CSCI 599: Information Integration on the Web**
University of Southern California, Spring 2002.
- **CSCI 541: Artificial Intelligence Planning**
Taught jointly with Dr. James Blythe, University of Southern California, Fall 1998.
- **CSCI 541: Artificial Intelligence Planning**
Taught jointly with Dr. Yolanda Gil, University of Southern California, Fall 1996.
- **CSCI 599: AI Planning Systems**
Taught jointly with Dr. Steve Chien, University of Southern California, Fall 1992.

Honors and Awards

- **Senior Member**, National Academy of Inventors (NAI), 2026.
- **Runner Up for Best In-Use Paper**, 2025.
Exploiting LLMs and Semantic Technologies to Build a Knowledge Graph of Historical Mining Data, Craig Knoblock, Binh Vu, Basel Shbita, Yao-Yi Chiang, Xiao Lin, Goran Muric, Pothula Punith Krishna, Jiyoung Pyo, Adriana Trejo-Sheu, and Meng Ye, The 24th International Semantic Web Conference (ISWC 2025).
- **Best Short Paper Award**, 2025.
Exploiting Polygon Metadata to Colorize Draft Maps, Fandel Lin, Craig Knoblock, Binh Vu, Basel Shbita, and Yao-Yi Chiang, The 33rd ACM SIGSPATIAL International conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2025).
- **Best Paper Honourable Mention**, 2025.
LDTR: Linear Object Detection Transformer for Accurate Graph Generation by Learning the N-hop Connectivity Information, Weiwei Duan, Yao-Yi Chiang, and Craig A. Knoblock, The 19th International Conference on Document Analysis and Recognition (ICDAR 2025).
- **First Place in the Map Feature Extraction Challenge**.
Jointly led a team of USC and the University of Minnesota in the DARPA AI for Critical Mineral Assessment Challenge, 2022.
- **AAIA Fellow**, Asia-Pacific Artificial Intelligence Association (AAIA), 2021.
- **IEEE Fellow**, Institute of Electrical and Electronics Engineers (IEEE), 2020.
- **Outstanding Engineering Merit Award**, Orange County Engineering Council, 2020.

- **Use-Inspired Research Award**, USC Viterbi School of Engineering, 2018.
- **Donald E. Walker Distinguished Service Award**, IJCAI, 2018.
- **ACM Fellow**, Association of Computing Machinery (ACM), 2017.
- **IAPR Best Paper Prize**, 2017.
Extracting Human Settlement Footprint from Historical Topographic Map Series Using Context-Based Machine Learning,
Johannes H. Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan, and Craig A. Knoblock,
8th International Conference on Pattern Recognition Systems (ICPRS-17).
- **Best Research Paper Award**, 2016.
Unsupervised entity resolution on multi-type graphs,
Linhong Zhu, Majid Ghasemi-Gol, Pedro Szekely, Aram Galstyan, and Craig A. Knoblock,
International Semantic Web Conference (ISWC).
- **Best In-Use Paper Award**, 2015.
Building and using a knowledge graph to combat human trafficking,
Pedro Szekely, Craig A. Knoblock, et al.,
14th International Semantic Web Conference (ISWC 2015).
- **Robert S. Engelmore Memorial Lecture Award**, 2014.
- **Finalist, 2nd LinkedUp Competition**, 2014.
LODStories: Learning About Art by Building Multimedia Stories,
Jianliang Chen, Yuting Liu, Dipanwita Maulik, Linda Xu, Hao Zhang, Craig A. Knoblock, Pedro Szekely and Miel Vander Sande.
- **Best In-Use Paper Award**, 2013.
Connecting the Smithsonian American Art Museum to the linked data cloud,
Pedro Szekely, Craig A. Knoblock, Fengyu Yang, Xuming Zhu, Eleanor Fink, Rachel Allen, and Georgina Goodlander,
Extended Semantic Web Conference (ESWC).
- **Best Research Paper Award**, 2012.
Discovering concept coverings in ontologies of linked data sources,
Rahul Parundekar, Craig A. Knoblock, and Jose Luis Ambite,
International Semantic Web Conference (ISWC).
- **Distinguished Scientist**, Association of Computing Machinery (ACM), 2008.
- **Fellow**, Association for the Advancement of Artificial Intelligence (AAAI), 2004.
- **Best Paper Award**,
Shou de Lin and Craig Knoblock,
Exploiting a search engine to develop more flexible web agents,
2003 IEEE/WIC International Conference on Web Intelligence (WI2003).
- **Meritorious Service Award**, Information Sciences Institute, 1999.
- **Teaching Fellowship**, Information Sciences Institute, Fall, 1996.
- **Best Paper Award**,
Evaluating the tradeoffs in partial-order planning algorithms,
Craig A. Knoblock and Qiang Yang,
Canadian Artificial Intelligence Conference, 1994.
- **Graduate Fellowship**, Air Force Laboratory, 1987-1991.
- **Class Scholar**, Syracuse University, 1984.
- **Upper Division Honors**, Syracuse University, 1984.
- **Lower Division Honors**, Syracuse University, 1982.
- **Regents Scholarship**, New York State, 1980-1984.

Invited Talks

- **From Papers to Knowledge: Building Knowledge Graphs from Technical Literature**
Keynote talk at The 8th IEEE International Conference on Cognitive Machine Intelligence, November 6, 2026.
- **From Tables to Knowledge Graphs: A Domain-Independent Approach to Semantic Table Interpretation**
Keynote talk at the The 17th Asian Conference on Machine Learning, December 9, 2025.
- **Mining the Past with AI: Building the World's Knowledge Graph of Critical Minerals**
Keynote talk at the The 30th International Conference on Technologies and Applications of AI, December 13, 2025.
- **From Developing New Drugs to Locating Critical Minerals: Using AI to Create Knowledge Graphs that Turn Data into Knowledge**
Keynote talk at the Proceedings of the 15th International SWAT4HCLS Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, February 28, 2024.
- **Exploiting Semantics for Integrating Data on Critical Minerals**
Invited talk at the Earth Science Information Partners (ESIP) Semantic Technologies Webinar Series, January 9, 2024.
- **Exploiting the Semantics of Tables to Clean and Align Data to Build Knowledge Graphs**
Invited talk at the IBM Tabular Data, Semantics and AI Seminar Series, January 26, 2023.
Keynote talk at the Workshop on Graph Learning for Industrial Applications: Finance, Crime Detection, Medicine and Social Media, NeurIPS 2022, December 9, 2022.
- **Building Spatio-temporal Knowledge Graphs from Historical Maps**
Keynote talk at the 1st ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs (GeoKG 2022), November 1, 2022.
- **Building and Using Knowledge Graphs to Turn Data into Knowledge**
Keynote talk at the Computing Conference, July 14, 2022. <https://youtu.be/lmbXVrySfvg>
Keynote talk at the Euro-Asia Conference on Frontiers of Computer Science and Information Technology, December 19, 2022
- **Acquiring and Exploiting the Semantics of Data**
Invited talk at the Open Data Science Conference (ODSC) West, Online & San Francisco, CA, November 18, 2021.
- **Building Knowledge Graphs to Solve Societal Problems: Applications, Challenges, and Solutions**
Keynote talk at the International Conference on Computing and Data Science, Online, August 2, 2020
- **Building Knowledge Graphs from Online Sources to Solve Societal Problems**
Keynote talk at The Web Conference Workshop on Linked Data on the Web and its Relationship with Distributed Ledgers (LDOW/LDDL), San Francisco, CA, May 13, 2019
- **Building Knowledge Graphs for Cyber Security**
Plenary talk at the AAAI Workshop on Artificial Intelligence for Cyber Security, Honolulu, Hawaii, January 27, 2019
- **Mining Data from the Deep and Dark Web to Combat the Illicit Art Trade**
Plenary talk at EuroMed 2018: International Conference on Digital Heritage, Nicosia, Cyprus, October 29, 2018
- **From Artwork to Cyber Attacks: Lessons Learned in Building Knowledge Graphs using Semantic Web Technologies**
Plenary talk at the U.S. Semantic Technologies Symposium, March 1, 2018
Invited talk at the University of Trento, Trento, Italy, April 18, 2018
- **Learning to Adapt to Sensor Changes and Failures**
Plenary talk at the IJCAI Workshop on AI for Internet of Things, August 19, 2017

- **Extracting, Aligning, and Linking Data to Build Knowledge Graphs**
Plenary talk at the 1st International Workshop on Industrial Knowledge Graphs, June 25, 2017
- **Lessons Learned in Building Human-Aware Systems for Data Science**
Invited talk at the AAAI Workshop on Human-Aware AI, February 4, 2017
- **Aligning and Integrating Data for Building Knowledge Graphs**
IC Academic Research Symposium, September 22, 2016
- **Creating Linked Knowledge for Cities**
AAAI Workshop on AI for Cities, January 25, 2015
- **Creating and Using Linked Knowledge**
Robert S. Engelmere Memorial Lecture Award, IAAI, July 29, 2014
AI2, Seattle, Wa, July 2, 2014
University of Texas at Austin, June 10, 2014
- **Data Integration for Big Data**
Huawei Strategy and Technology Workshop, May 13, 2013
- **The Role of Semantics in Big Data Integration**
2013 Semantic Technology & Business Conference, June 4, 2013
- **Learning Semantics of Sources for Big Data Integration**
Invited talk at the University of Nebraska at Lincoln, March 27, 2013
Invited talk at the University of Nebraska at Omaha, March 28, 2013
- **Challenges to the Success of Linked Open Data** Invited talk at the Planet Data Roadmapping Workshop at the 2012 Extended Semantic Web Conference, May 30, 2012.
- **Harvesting Named Geographic Features from Raster Maps**
Invited talk at GeoINT 2011 Symposium Interoperability Tech Talks, San Antonio, TX, Oct 18, 2011.
- **Interactive Information Integration for Network Extraction, Discovery, and Analysis**
Invited talk at the Technology Visions for Network Extraction, Discovery, and Analysis Workshop, Vernon, NY, Sept 21, 2011.
- **A General Approach to Discovering, Registering, and Extracting Features from Raster Maps**
Invited talk at the University of California, Santa Barbara, CA, February 3, 2011
Invited talk at Syracuse University, Syracuse, NY, October 6, 2010
Invited talk at Google, Inc., Mountain View, CA, August 25, 2010
- **A Reference-Set Approach to Information Extraction from Unstructured, Ungrammatical Data Sources**
Invited talk at the University of Washington, Seattle, WA, January 21, 2011
Invited talk at the University of Lugano, Lugano, Switzerland, July 12, 2010
- **Building Semantic Descriptions of Linked Data**
Invited talk at the 1st International Workshop on Consuming Linked Data (COLD 2010) at the International Semantic Web Conference, 2010.
Shanghai, China, November 8, 2010
- **Interactively Building Geospatial Mashups**
Keynote talk at the 2nd International Workshop on New Trends in Information Integration
Long Beach, CA, March 5, 2010
- **A General Approach to Discovering, Registering, and Extracting Features from Raster Maps**
Invited talk at the Conference on Document Recognition and Retrieval of the IS&T/SPIE 19th Annual Symposium International Symposium on Electronic Imaging
San Jose, CA, January 20, 2010

- **Geospatial Data Integration for High-level Fusion**
Invited talk at the ONR Workshop Panel on Directions for Higher-Level Fusion Research: Needs and Capabilities
Seattle, WA, July 6, 2009
- **Discovering and Building Semantic Models of Web Sources**
Keynote address at 6th Annual European Semantic Web Conference (ESWC2009)
Heraklion, Greece, June 3, 2009
- **Web-based Learning**
Invited talk at the AFRL Next Generation Reasoning and Learning: Theory and Applications Workshop
Vernon, NY, May 27, 2009
- **Exploiting Open Source Data for Imagery Understanding**
Invited talk at NRO Tech Seminar
Chantilly, VA, February 20, 2009
- **Automatically Discovering, Extracting and Modeling Web Sources for Information Integration**
Invited talk at Università degli Studi di Genova
Genoa, Italy, December 15, 2008
- **Building Mashups by Example**
Invited talk at Microsoft Research
Bellevue, WA, August 1, 2008
- **Finding, Extracting, and Integrating Data from Maps**
Invited talk at University of Nebraska
Lincoln CSE Colloquium Series, November 29, 2007.
- **Beyond the Elves: Making Intelligent Agents Intelligent**
Invited talk at AAAI Spring Symposium on What Went Wrong and Why
Palo Alto, CA, March 28, 2006
- **Learning to Optimize Plan Execution in Information Agents**
Plenary talk at the 6th International Conference on Case-Based Reasoning (ICCBR 2005)
Chicago, IL, USA, August 25, 2005
- **Mining Data from Online and Geospatial Information Sources**
Invited talk at National Geospatial Intelligence Agency (Oct 19, 2004), Central Intelligence Agency (Oct 19, 2004), National Security Agency (Oct 20, 2004)
- **Building Software Agents for Planning, Monitoring, and Optimizing Travel**
Plenary talk at the Eleventh International Conference On Information Technology and Travel & Tourism
Cairo, Egypt, January, 26, 2004
- **Deploying Information Agents on the Web**
Plenary talk at the International Joint Conference on Artificial Intelligence (IJCAI-03)
Acapulco, Mexico, August 15, 2003
- **Integrating Online and Geospatial Information Sources**
Plenary talk at the Summer Assembly of the University Consortium for Geographic Information Science
Monterey, CA, June 17, 2003
- **Information Integration on the Web**
Invited talk at the Planet 2002 Summer School on Planning on the Web, September 2002
- **Exploring Real World Planning**
Invited talk at the AIPS 2002 Workshop, April, 2002
- **Electronic Commerce in the Information Industries**
Invited talk at The 36th Annual Graduate School of Library and Information Science Clinic, April 2-4, 2000

- **Abstraction and Hierarchy in Reinforcement Learning**
Invited talk at NIPS-98 workshop, 1998

Professional Activities

Elected Positions

- **Past President**, International Joint Conferences on Artificial Intelligence, 2013-2015.
- **President**, International Joint Conferences on Artificial Intelligence, 2011-2013.
- **Trustee**, International Joint Conference on Artificial Intelligence, 2007-2015.
- **President**, International Conference on Automated Planning and Scheduling, 2006-2008.
- **President-Elect**, International Conference on Automated Planning and Scheduling, 2004-2006.
- **Treasurer**, Executive Council of the International Conference on Automated Planning and Scheduling, 2002-2004.
- **Executive Council Member**, AAAI, 2001-2004.

2026

- **Program Committee, Research Track**, International Semantic Web Conference, 2026
- **Program Committee, Poster and Demo Track**, Semantics, 2026
- **Program Committee**, 6th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment (Sci-K), 2026
- **Steering Committee**, IEEE Transactions on Artificial Intelligence, 2026
- **Associate Editor**, Spatial Algorithms and Systems, 2026
- **Editorial Board**, Data Journal, 2026
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2026
- **Publication Committee**, AAAI Press, 2026

2025

- **Program Committee, In-Use Track**, International Semantic Web Conference, 2025
- **Program Committee, Poster and Demo Track**, International Semantic Web Conference, 2025
- **Steering Committee**, IEEE Transactions on Artificial Intelligence, 2025
- **Associate Editor**, Spatial Algorithms and Systems, 2024
- **Editorial Board**, Data Journal, 2025
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2025
- **Publication Committee**, AAAI Press, 2025

2024

- **Program Committee, In-Use Track**, International Semantic Web Conference, 2024
- **Program Committee, Poster and Demo Track**, International Semantic Web Conference, 2024
- **Steering Committee**, IEEE Transactions on Artificial Intelligence, 2024
- **Associate Editor**, Spatial Algorithms and Systems, 2024

- **Editorial Board**, Data Journal, 2024
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2024
- **Publication Committee**, AAAI Press, 2024

2023

- **Program Committee, Research Track**, International Semantic Web Conference, 2023
- **Program Committee, In-Use Track**, International Semantic Web Conference, 2023
- **Steering Committee**, IEEE Transactions on Artificial Intelligence, 2023
- **Associate Editor**, Spatial Algorithms and Systems, 2023
- **Editorial Board**, Data Journal, 2023
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2023
- **Publication Committee**, AAAI Press, 2023

2022

- **Program Committee, Research Track**, International Semantic Web Conference, 2022
- **Program Committee, In-Use Track**, International Semantic Web Conference, 2022
- **Program Committee, Poster and Demos Track**, Extended Semantic Web Conference, 2022
- **Steering Committee**, IEEE Transactions on Artificial Intelligence, 2022
- **Associate Editor**, Spatial Algorithms and Systems, 2022
- **Editorial Board**, Data Journal, 2022
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2022
- **Publication Committee**, AAAI Press, 2022
- **Finance Committee**, IJCAI Inc., 2022

2021

- **Co-Organizer**, IJCAI Workshop on Applied Semantic Extraction and Analytics (ASEA 2021), 2021
- **Program Committee**, IJCKG, 2021
- **Program Committee, In-Use Track**, International Semantic Web Conference, 2021
- **Steering Committee**, IEEE Transactions on Artificial Intelligence, 2021
- **Associate Editor**, Spatial Algorithms and Systems, 2021
- **Editorial Board**, Data Journal, 2021
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2021
- **Publication Committee**, AAAI Press, 2021
- **Finance Committee**, IJCAI Inc., 2021

2020

- **Area Chair**, AAAI 2021
- **Program Committee**, 24th European Conference on Artificial Intelligence (ECAI 2020), 2020
- **Steering Committee**, IEEE Transactions on Artificial Intelligence, 2020
- **Associate Editor**, Spatial Algorithms and Systems, 2020
- **Editorial Board**, Data Journal, 2020
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2020
- **Publication Committee**, AAAI Press, 2020
- **Finance Committee**, IJCAI Inc., 2020

2019

- **Program Committee, Journal Track**, International Semantic Web Conference, 2019
- **Program Committee, In-Use Track**, International Semantic Web Conference, 2019
- **Program Committee**, WWW 2019 Workshop on Knowledge Graph Technology and Applications
- **Associate Editor**, Spatial Algorithms and Systems, 2019
- **Editorial Board**, Data Journal, 2019
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2019
- **Publication Committee**, AAAI Press, 2019
- **Finance Committee**, IJCAI Inc., 2019

2018

- **Organizer**, Workshop on Knowledge Base Construction, Reasoning and Mining, 2018
- **Program Committee, In-Use Track**, International Semantic Web Conference, 2018
- **Program Committee**, 2nd Workshop on Enabling Open Semantic Science, 2018
- **Associate Editor**, Spatial Algorithms and Systems, 2018
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2018
- **Publication Committee**, AAAI Press, 2018
- **Finance Committee**, IJCAI Inc., 2018

2017

- **Senior Program Committee**, AAAI 2017.
- **Senior Program Committee, Research Track**, International Semantic Web Conference, 2017
- **Program Committee, In-Use Track**, International Semantic Web Conference, 2017
- **Program Committee**, Extended Semantic Web Conference, 2017
- **Program Committee**, WWW 2017 Special Track on Semantics and Knowledge, 2017
- **Program Committee**, 1st International Workshop on Industrial Knowledge Graphs, 2017
- **Program Committee**, 1st Workshop on Enabling Open Semantic Science, 2017
- **Associate Editor**, Spatial Algorithms and Systems, 2017
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2017
- **Publication Committee**, AAAI Press, 2017
- **Finance Committee**, IJCAI Inc., 2017

2016

- **Senior Program Committee**, IJCAI 2016
- **Senior Program Committee**, AAAI 2016
- **Program Committee**, Extended Semantic Web Conference, 2016
- **Program Committee**, International Semantic Web Conference, 2016
- **Program Committee**, 15th International Semantic Web Conference Applications Track
- **Program Co-Chair**, 5th Workshop on Knowledge Discovery and Data Mining meets Linked Open Data, 2016
- **Program Committee**, Workshop on Linked Data for Information Extraction, 2016.
- **Associate Editor**, Artificial Intelligence, 2009-2016
- **Associate Editor**, Spatial Algorithms and Systems, 2015-2016
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2016
- **Publication Committee**, AAAI Press, 2016.
- **Finance Committee**, IJCAI Inc., 2016

2015

- **Senior Program Committee**, Conference on Artificial Intelligence, AI & the Web Track, 2015
- **Program Committee**, Workshop on Linked Data for Information Extraction, 2015.
- **Program Committee**, 11th International Conference on Data Integration in the Life Sciences, 2015.
- **Program Committee**, ISWC 5th International Workshop on Linked Science 2015.
- **Associate Editor**, Artificial Intelligence, 2009-2016
- **Associate Editor**, Spatial Algorithms and Systems, 2015-2016
- **Editorial Board Member**, Journal of Web Semantics, 2010-2015.
- **Advisory Board Member**, ACM Transactions on Intelligent Systems and Technology, 2015.
- **Editorial Board Member**, AAAI Press, 2015.

2014

- **Program Chair**, International Semantic Web Conference, In-Use Track, 2014
- **Program Committee**, Conference on Artificial Intelligence (AAAI), 2014
- **Program Committee**, AAAI Workshop on Discovery Informatics: Scientific Discoveries Enabled by AI, 2014.
- **Program Committee**, Workshop on Linked Data for Information Extraction, 2014. .
- **Student Travel Award Committee**, ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2014.
- **Associate Editor**, *Artificial Intelligence*, 2009-2015
- **Editorial Board Member**, *Journal of Web Semantics*, 2010-2015.
- **Advisory Board Member**, *ACM Transactions on Intelligent Systems and Technology*, 2014.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2014.
- **Editorial Board Member**, *AAAI Press*, 2014.

2013

- **Conference Co-Chair**, ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2013.
- **Senior Program Committee Member**, Conference on Artificial Intelligence, AI & the Web Track, 2013
- **Program Committee**, AAAI 2013 Workshop on Semantic Cities.
- **Organizing Committee**, Tenth International Workshop on Information Integration on the Web (IIWeb 2013).
- **Program Committee**, 4th International Workshop on Consuming Linked Data, 2013.
- **Program Committee**, Extended Semantic Web Conference, Tutorials, 2013.
- **Program Committee**, Workshop on Linked Data for Information Extraction, 2013.
- **Program Committee**, ISWC 3rd Workshop on Linked Science 2013
- **Program Committee**, Semantics For Big Data, AAAI Fall Symposium 2013
- **Associate Editor**, *Artificial Intelligence*, 2009-2015
- **Editorial Board Member**, *Journal of Web Semantics*, 2010-2013.
- **Advisory Board Member**, *ACM Transactions on Intelligent Systems and Technology*, 2013.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2013.
- **Editorial Board Member**, *AAAI Press*, 2013.

2012

- **Conference Co-Chair**, ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2012.
- **Workshop Chair**, Workshop at the Extended Semantic Web Conference on Linked APIs, 2012.
- **Area Chair**, Twenty-Sixth Conference on Artificial Intelligence, 2012.
- **Organizing Committee**, Computing Community Consortium Workshop on Spatial Computing, 2012
- **Steering Committee**, AAAI 2012 Workshop on Semantic Cities.
- **Organizing Committee**, Ninth International Workshop on Information Integration on the Web (IIWeb 2012).
- **Program Committee**, Extended Semantic Web Conference, 2012.
- **Program Committee**, International Semantic Web Conference, 2012.
- **Program Committee**, Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data, 2012.
- **Associate Editor**, *Artificial Intelligence*, 2009-2012
- **Editorial Board Member**, *Journal of Web Semantics*, 2010-2013.
- **Advisory Board Member**, *ACM Transactions on Intelligent Systems and Technology*, 2012.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2012.
- **Editorial Board Member**, *AAAI Press*, 2012.

2011

- **Conference Chair**, International Joint Conference on Artificial Intelligence, 2011
- **Associate Editor**, *Artificial Intelligence*, 2009-2012
- **Editorial Board Member**, *Journal of Web Semantics*, 2010-2013.
- **Advisory Board Member**, *ACM Transactions on Intelligent Systems and Technology*, 2011.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2011.
- **Editorial Board Member**, *AAAI Press*, 2011.
- **Special Issue Editor**, *Journal of Web Semantics* on Dealing with the Messiness of the Web of Data, 2011.
- **Program Committee Member**, ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2011.
- **Steering Committee**, IIWeb 2011 : Eighth International Workshop on Information Integration on the Web, 2011.

2010

- **Conference Chair**, International Joint Conference on Artificial Intelligence, 2011
- **Associate Editor**, *Artificial Intelligence*, 2009-2012
- **Advisory Board Member**, *ACM Transactions on Intelligent Systems and Technology*, 2010.
- **Editorial Board Member**, *Journal of Web Semantics*, 2010-2013.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2010.
- **Editorial Board Member**, *AAAI Press*, 2010.
- **Senior Program Committee Member**, Twenty-Fourth Conference on Artificial Intelligence, AI & the Web Special Track, 2010.
- **Program Committee Member**, ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010.

2009

- **Conference Chair**, International Joint Conference on Artificial Intelligence, 2011
- **Co-organizer**, Workshop on Information Integration on the Web, 2009.
- **Associate Editor**, *Artificial Intelligence*, 2009-2012
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2009.
- **Editorial Board Member**, *AAAI Press*, 2009.
- **Senior Program Committee Member**, Conference on Intelligent User Interfaces, 2009.
- **Program Committee Member**, Twenty-Third Conference on Artificial Intelligence, AI & the Web Special Track, 2009.
- **Program Committee Member**, International Conference on GeoSpatial Semantics, 2009.
- **Program Committee Member**, ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2009.

2008

- **Conference Chair**, International Joint Conference on Artificial Intelligence, 2011
- **Conference Co-Chair**, Twenty-Second Conference on Artificial Intelligence, AI & the Web Special Track, 2008.
- **Vice Chair**, International Semantic Web Conference, 2008.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2008.
- **Editorial Board Member**, *AAAI Press*, 2008.
- **Senior Program Committee**, Conference on Intelligent User Interfaces, 2008.
- **Program Committee Member**, International Conference on Advances in Geographic Information Systems, 2008.
- **Program Committee Member**, Second Workshop on Analytics for Noisy Unstructured Text Data, 2008
- **Panel Member**, National Geospatial-Intelligence Agency Multi-Source, Multi-INT Fusion Panel, State College, PA, December 2-3, 2008.

2007

- **Conference Chair**, International Joint Conference on Artificial Intelligence, 2011
- **Senior Program Committee Member**, National Conference on Artificial Intelligence, 2007.
- **Editorial Board Member**, *Computational Intelligence*, 2005-2007.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2007.
- **Editorial Board Member**, *AAAI Press*, 2005-2007.
- **Program Committee Member**, Workshop on Information Integration on the Web, 2007.

2006

- **Senior Program Committee Member**, International Joint Conference on Artificial Intelligence, 2007.
- **Senior Program Committee Member**, National Conference on Artificial Intelligence, 2006.
- **Co-organizer**, IJCAI 2007 Workshop on Analytics for Noisy Unstructured Text Data
- **Review Panel Member**, Marine Geosciences External Review of NRL's Battlespace Environments Technology S&T Program, Stennis, September 19-20, 2006.
- **Editorial Board Member**, *Computational Intelligence*, 2005-2007.
- **Editorial Board Member**, *Journal on Foundations and Trends in Web Science*, 2006.
- **Editorial Board Member**, *AAAI Press*, 2005-2007.
- **Program Committee Member**, Workshop on Information Integration on the Web, 2006.

2005

- **Participant**. National Academy of Sciences Workshop on Using Information Technology to Enhance Disaster Management, Washington, D.C., June 22-23, 2005.
- **Editorial Board Member**, *Computational Intelligence*, 2005-2007.
- **Founding Member**, Editorial Board for the *Journal on Foundations and Trends in Web Science*, 2005.

- **Editorial Board Member**, *AAAI Press*, 2005-2007.
- **Program Committee Member**, ACM Conference on Geospatial Information Science (ACM GIS), 2005.
- **Program Committee Member**, International Conference on Geospatial Semantics (GeoS), 2005.
- **Program Committee Member**, Workshop on Representation and Analysis of Web Space (RAWS-05), 2005.
- **Program Committee Member**, International Joint Conference on Artificial Intelligence, 2005.
- **Program Committee Member**, National Conference on Artificial Intelligence, 2005.
- **Program Committee Member**, International Conference on Automated Planning and Scheduling, 2005.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 2005.

2004

- **Guest Editor**, Special issue of *IEEE Intelligent Systems* on Mining the Web, 19((6), November/December, 2004.
- **Senior Program Committee Member**, Semantic Web Conference, 2004
- **Senior Program Committee Member**, National Conference on Artificial Intelligence, 2004.
- **Program Committee Member**, International Conference on Automated Planning and Scheduling, 2004.
- **Program Committee Member**, 17th International FLAIRS Conference Special Track on AI and the Web, 2004.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 2004.

2003

- **Guest Editor**, Special issue of *IEEE Intelligent Systems* on Information Integration on the Web, September, 2003.
- **Co-Chair**, IJCAI'03 Workshop on Information Integration on the Web, 2003.
- **Program Committee Member**, International Conference on Service Oriented Computing, 2003.
- **Program Committee Member**, KDD'03 Workshop on Data Cleaning, Record Linkage and Object Consolidation, 2003.
- **Program Committee Member**, ISWC'03 Semantic Integration Workshop.
- **Program Committee Member**, Conference on AI Planning and Scheduling, 2003.
- **Program Committee Member**, Conference on Innovative Applications of Artificial Intelligence, 2003
- **Editorial Board Member**, *IEEE Intelligent Systems*, 2003.

2002

- **Member**, Joint European-US Working Group on Information Extraction from Digital Libraries, 2002.
- **Member**, DARPA IXO panel study on Information Integration, 2002.
- **Panel Member**, Robotic, Intelligent and Information Systems Evaluation, Science and Technology Foundation, Ministry of Science and Technology, Portugal, February, 2002.

- **Program Committee Member**, Conference on Innovative Applications of Artificial Intelligence, 2002.
- **Program Committee Member**, Conference on AI Planning and Scheduling, 2002.
- **Program Committee Member**, National Conference on Artificial Intelligence, 2002.
- **Program Committee Member**, AIPS 2002 Workshop on Exploring Real-World Planning, 2002.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 2002.

2001

- **Co-Chair**, NASA RIACS Panel on Information Management Requirements, 2001-2002.
- **Program Committee Member**, Conference on Innovation Applications of Artificial Intelligence, 2001.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 2001.

2000

- **Program Co-Chair**, International Conference on AI Planning and Scheduling, 2000.
- **Senior Program Committee Member**, National Conference on Artificial Intelligence, 2000.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 2000.

1999

- **Review Board Member**, Jet Propulsion Laboratory Annual Technology Program Review of the Telecommunications and Mission Operations Directorate (TMOD), 1999.
- **Reviewer**, International Joint Conference on Artificial Intelligence, 1999.
- **Publicity Chair**, Autonomous Agents, 1999.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 1999.

1998

- **Senior Program Committee Member**, National Conference on Artificial Intelligence, 1998.
- **Co-organizer**, AAAI Workshop on AI and Information Integration at the Fourteenth National Conference on Artificial Intelligence, 1998.
- **Program Committee Member**, International Conference on Artificial Intelligence Planning Systems, 1998.
- **Program Committee Member**, International Conference on Autonomous Agents, 1998.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 1998.

1997

- **Senior Program Committee Member**, National Conference on Artificial Intelligence, 1997.
- **Editor**, Department on Trends and Controversies for *IEEE Expert*, 1997.
- **Program Committee Member**, International Conference on Autonomous Agents, 1997.
- **Reviewer**, European Conference on Planning, 1997.
- **Editorial Board Member**, *IEEE Intelligent Systems*, 1997.
- **Editorial Board Member**, *Journal of Artificial Intelligence Research*, 1997.

1996

- **Editorial Board Member**, *Journal of Artificial Intelligence Research*, 1996.
- **Editor**, Department on Trends and Controversies for *IEEE Expert*, 1996.
- **Program Committee Member**, National Conference on Artificial Intelligence, 1996.
- **Reviewer**, International Conference on Artificial Intelligence Planning Systems, 1996.

1995

- **Editorial Board Member**, *Journal of Artificial Intelligence Research*, 1995.
- **Program Committee Member**, Conference on Artificial Intelligence Applications, 1995.
- **Program Committee Member**, Conference on Cooperative Information Systems, 1995.
- **Reviewer**, International Joint Conference on Artificial Intelligence, 1995.
- **Co-organizer**, AAAI Spring Symposium on Information Gathering from Distributed, Heterogeneous Environments, 1995.
- **Panel Member**, National Science Foundation proposal review panel, 1995.

1993

- **Reviewer**, International Joint Conference on Artificial Intelligence, 1993.

1992

- **Associate Chair**, Tenth National Conference on Artificial Intelligence, 1992.
- **Program Committee Member**, Conference on Principles of Knowledge Representation and Reasoning, 1992.

Journal Reviewer

- *Artificial Intelligence*
- *Constraints, An International Journal*
- *Machine Learning*
- *Journal of Artificial Intelligence Research*
- *Journal of the ACM*
- *Journal of Experimental and Theoretical Artificial Intelligence*
- *Journal of Web Semantics*
- *IEEE Transactions on Knowledge and Data Engineering*
- *IEEE Transactions on Pattern Analysis and Machine Intelligence*
- *IEEE Intelligent Systems*
- *International Journal of Cooperative Information Systems*
- *Theory and Practice of Object Systems*
- *World Wide Web Journal*

Graduated Students

- **Fandel Lin, 2026**
Learning to Interpret Historical Maps by Exploiting Polygon Metadata
Ph.d., Department of Computer Science, University of Southern California, 2026
- **Binh Le Thanh Vu, 2024**
Exploiting Web Tables and Knowledge Graphs for Creating Semantic Descriptions of Data Sources
Ph.d., Department of Computer Science, University of Southern California, 2024
- **Basel Sbhita, 2024**
Transforming Unstructured Historical and Geographic Data into Spatio-Temporal Knowledge Graphs
Ph.d., Department of Computer Science, University of Southern California, 2024
- **Minh Tran Xuan Pham, 2022**
Robust and Proactive Error Detection and Correction in Tables
Ph.d., Department of Computer Science, University of Southern California, 2022
- **Marianna Bucchi, 2019**
Exploiting the Semantic Web for the Automatic Extraction of Los Angeles City Data
M.S., Management Engineering, Politecnico di Milano
- **Yuan Shi, 2019**
Learning to Adapt to Sensor Changes and Failures
Ph.d., Department of Computer Science, University of Southern California, 2019
- **Mohsen Taheriyani, 2015**
Learning the Semantics of Structured Data Sources
Ph.d., Department of Computer Science, University of Southern California, 2015
Current position: Research Scientist, Google
- **Bo Wu, 2015**
Iteratively Learning Data Transformation Programs from Examples
Ph.d., Department of Computer Science, University of Southern California, 2015
Current position: Yelp.com
- **Daniel W. Goldberg, 2010**
Spatial Approaches to Reducing Error in Geocoded Data
Ph.D., Department of Computer Science, University of Southern California, 2010
Jointly advised with Professor John Wilson
Current position: Assistant Professor, Texas A&M University, College Station
- **Yao-Yi Chiang, 2010**
Harvesting Geographic Features from Heterogeneous Raster Maps
Ph.D., Department of Computer Science, University of Southern California, 2010
Current position: Assistant Professor, Research, Spatial Sciences, University of Southern California
- **Matthew Michelson, 2009**
A Reference-set Approach to Information Extraction from Unstructured, Ungrammatical Data Sources
Ph.D., Department of Computer Science, University of Southern California, 2009
Current position: Researcher, Inferlink, Inc.
- **Martin Michalowski, 2008**
A General Approach to Using Problem Instance Data for Model Refinement in Constraint Satisfaction Problems
Ph.D., Department of Computer Science, University of Southern California, 2008
Current position: Assistant Professor, University of Minnesota
- **Rattapoom Tuchinda, 2008**
Building Mashups by Example
Ph.D., Department of Computer Science, University of Southern California, 2008
Current position: National Electronics and Computer Technology Center, Thailand

- **Snehal Thakkar, 2007**
A Data Integration Approach to Dynamically Fusing Geospatial Sources
Ph.D., Department of Computer Science, University of Southern California, 2005
Current position: Google, Inc.
- **Mark James Carman, 2006**
Learning Semantic Definitions of Information Sources on the Internet
Ph.D. Thesis, Department of Information and Communication Technology, University of Trento, 2006
Current position: Associate Professor at Politecnico di Milano
- **Ching-Chien Chen, 2005**
Automatically and Accurately Conflating Road Vector Data, Street Maps and Orthoimagery
Ph.D., Department of Computer Science, University of Southern California, 2005
Jointly advised with Professor Cyrus Shahabi
Current position: Director of Research and Development, TerraGo Technologies
- **Matthew Michelson, 2005**
Building Queryable Datasets from Ungrammatical and Unstructured Sources
M.S., Department of Computer Science, University of Southern California, 2005
Current position: Researcher, InferLink
- **Rahul Bakshi, 2004**
Integrating and reasoning about online sources to accurately geocode addresses
M.S., Department of Computer Science, University of Southern California, 2004
Current position: Software Engineer, SiRF Technology
- **Greg Barish, 2003**
Speculative Plan execution for information agents
Ph.D., Department of Computer Science, University of Southern California, 2003
Current position: Chief Technology Officer, InferLink
- **Sheila Tejada, 2002**
Learning object identification rules for information integration
Ph.D., Department of Computer Science, University of Southern California, 2002
Current position: Lecturer, University of Southern California
- **Ion Alexandru Muslea, 2002**
Active Learning with Multiple Views
Ph.D., Department of Computer Science, University of Southern California, 2002
Current position: Research Scientist, Amazon
- **Naveen Ashish, 2000**
Optimizing information mediators by selectively materializing data
Ph.D., Department of Computer Science, University of Southern California, 2000
Current Position: Research Scientist, InferLink
- **Jose Luis Ambite, 1998**
Planning by rewriting
Ph.D., Department of Computer Science, University of Southern California, 1998
Current Position: Research Lead and Research Associate Professor, Information Sciences Institute, University of Southern California
- **Chun-Nan Hsu, 1996**
Learning effective and robust semantic knowledge for database query optimization
Ph.D., Department of Computer Science, University of Southern California, 1996
Current Position: Research Scientist, University of California at San Diego

Publications

Journal Articles

1. Fandel Lin, Han Zhang, T. K. Satish Kumar, and Craig A. Knoblock. An integrated approach to multi-agent scheduling with bounded objectives in an urban setting. *ACM Trans. Spatial Algorithms Syst.*, 12(1), January 2026.
2. Basel Shbita, Craig A Knoblock, Weiwei Duan, Yao-Yi Chiang, Johannes H Uhl, and Stefan Leyk. Building spatio-temporal knowledge graphs from vectorized topographic historical maps. *Semantic Web*, 14(3):527–549, 2023.
3. Yolanda Gil, Daniel Garijo, Deborah Khider, Craig A. Knoblock, Varun Ratnakar, Maximiliano Osorio, Hernan Vargas, Minh Pham, Jay Pujara, Basel Shbita, Binh Vu, Yao-Yi Chiang, Dan Feldman, Yijun Lin, Hayley Song, Vipin Kumar, Ankush Khandelwal, Michael Steinbach, Kshitij Tayal, Shaoming Xu, Suzanne A. Pierce, Lissa Pearson, Daniel Hardesty-Lewis, Ewa Deelman, Rafael Ferreira Da Silva, Rajiv Mayani, Armen R. Kemanian, Yuning Shi, Lorne Leonard, Scott Peckham, Maria Stoica, Kelly Cobourn, Zeya Zhang, Christopher Duffy, and Lele Shu. Artificial intelligence for modeling complex systems: Taming the complexity of expert models to improve decision making. *ACM Transactions on Interactive Intelligent Systems*, 11(2), July 2021.
4. Johannes H. Uhl, Stefan Leyk, Zekun Li, Weiwei Duan, Basel Shbita, Yao-Yi Chiang, and Craig A. Knoblock. Combining remote-sensing-derived data and historical maps for long-term back-casting of urban extents. *Remote Sensing*, 13(18), 2021.
5. Yuan Shi, Ang Li, T. K. Satish Kumar, and Craig A. Knoblock. Building survivable software systems by automatically adapting to sensor changes. *Applied Sciences*, 11(11), 2021.
6. Weiwei Duan, Yao-Yi Chiang, Stefan Leyk, Johannes H. Uhl, and Craig A. Knoblock. Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning. *International Journal of Geographical Information Science*, 34(4):824–849, 2020.
7. Avi Pfeffer, Curt Wu, Gerald Fry, Kenny Lu, Steve Marotta, Mike Reposa, Yuan Shi, TK Satish Kumar, Craig A Knoblock, David Parker, Irfan Muhammad, and Chris Novakovic. Software adaptation for an unmanned undersea vehicle. *IEEE Software*, 36(2):91–96, 2019.
8. Ying Zhang, Qunfei Ma, Yao-Yi Chiang, Craig Knoblock, Xin Zhang, Puhai Yang, Minghe Gao, and Xiang Hu. Extracting geographic features from the internet: A geographic information mining framework. *Knowledge-Based Systems*, 174:57–72, 2019.
9. Johannes H Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan, and Craig A Knoblock. Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks. *IEEE Access*, 2019.
10. Weiwei Duan, Yao-Yi Chiang, Stefan Leyk, Johannes H Uhl, and Craig A Knoblock. Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning. *International Journal of Geographical Information Science*, pages 1–26, 2019.
11. Johannes H Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan, and Craig A. Knoblock. Spatialising uncertainty in image segmentation using weakly supervised convolutional neural networks: a case study from historical map processing. *IET Image Processing*, 12(11):2084–2091, 2018.
12. Johannes H Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan, and Craig A. Knoblock. Map archive mining: Visual-analytical approaches to explore large historical map collections. *ISPRS International Journal of Geo-Information*, 7(4):148, 2018.
13. Mohsen Taheriyani, Craig A. Knoblock, Pedro Szekely, and Jose Luis Ambite. Learning the semantics of structured data sources. *Journal of Web Semantics*, 37, 2016.
14. Pedro Szekely, Craig A. Knoblock, Fengyu Yang, Xuming Zhu, Eleanor Fink, Rachel Allen, and Georgina Goodlander. Publishing the data of the smithsonian american art museum to the linked data cloud. *International Journal of Humanities and Art Computing (IJHAC)*, 8:152–166, 2014.

15. Yao-Yi Chiang, Stefan Leyk, and Craig A. Knoblock. A survey of digital map processing techniques. *ACM Computing Surveys*, 47(1):1–44, 2014.
16. Yao-Yi Chiang and Craig A. Knoblock. Recognizing text in raster maps. *Geoinformatica*, pages 1–27, 2014. Published Online.
17. Yao-Yi Chiang and Craig A. Knoblock. A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition*, 16(1):55–81, 2013.
18. Yao-Yi Chiang and Craig A. Knoblock. A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition*, 2011.
19. Rattapoom Tuchinda, Craig A. Knoblock, and Pedro Szekely. Building mashups by demonstration. *ACM Transactions on the Web (TWEB)*, 5(3), July 2011. <http://dx.doi.org/10.1145/1993053.1993058>.
20. Aman Goel, Matthew Michelson, and Craig A. Knoblock. Harvesting maps on the web. *International Journal on Document Analysis and Recognition*, 2010. <http://dx.doi.org/10.1007/s10032-010-0136-2>.
21. Matthew Michelson and Craig A. Knoblock. Constructing reference sets from unstructured, ungrammatical text. *Journal of Artificial Intelligence Research*, 38:189–221, 2010.
22. Daniel W. Goldberg, John P. Wilson, and Craig A. Knoblock. Extracting geographic features from the internet to automatically build detailed regional gazetteers. *International Journal of Geographic Information Science*, 2009.
23. Yao-Yi Chiang, Craig A. Knoblock, Cyrus Shahabi, and Ching-Chien Chen. Automatic and accurate extraction of road intersections from raster maps. *Geoinformatica*, 2008. <http://dx.doi.org/10.1007/s10707-008-0046-3>.
24. Ching-Chien Chen, Craig A. Knoblock, and Cyrus Shahabi. Automatically and accurately conflating raster maps with orthoimagery. *Geoinformatica*, 12(3):377–410, 2008. <http://dx.doi.org/10.1007/s10707-007-0033-0>.
25. Jim Blythe, Dipsy Kapoor, Craig A. Knoblock, Kristina Lerman, and Steven Minton. Information integration for the masses. *Journal of Universal Computer Science*, 14(11):1811–1837, 2008.
26. Daniel W. Goldberg, John P. Wilson, Craig A. Knoblock, Beate Ritz, and Myles G Cockburn. An effective and efficient approach for manually improving geocoded data. *International Journal of Health Geographics*, 7(60), 2008. doi:10.1186/1476-072X-7-60.
27. Matthew Michelson and Craig A. Knoblock. Creating relational data from unstructured and ungrammatical data sources. *Journal of Artificial Intelligence Research (JAIR)*, 31:543–590, 2008.
28. Greg Barish and Craig A. Knoblock. Speculative plan execution for information gathering. *Artificial Intelligence*, 172(4-5):413–453, 2008. <http://dx.doi.org/10.1016/j.artint.2007.08.002>.
29. Craig A. Knoblock, Jose Luis Ambite, Kavita Ganesan, Maria Muslea, Steven Minton, Greg Barish, Evan Gamble, Claude Nanjo, Kane See, Cyrus Shahabi, , and Ching-Chien Chen. Entitybases: Compiling, organizing and querying massive entity repositories. In *International Conference on Artificial Intelligence (ICAI’07)*, Las Vegas, Nevada, USA, June 2007.
30. Matthew Michelson and Craig A. Knoblock. Unsupervised information extraction from unstructured, ungrammatical data sources on the world wide web. *International Journal of Document Analysis and Recognition (IJDAR), Special Issue on Noisy Text Analytics*, 10(3-4):211–226, 2007.
31. Mark Carman and Craig A. Knoblock. Learning semantic definitions of online information sources. *Journal of Artificial Intelligence Research (JAIR)*, 30:1–50, 2007.
32. Kristina Lerman, Anon Plangprasopchok, and Craig A. Knoblock. Semantic labeling of online information sources. *International Journal on Semantic Web and Information Systems*, 3(3):36–56, 2007.
33. Daniel W. Goldberg, John P. Wilson, and Craig A. Knoblock. From text to geographic coordinates: The current state of geocoding. *Journal of the Urban and Regional Information Systems Association*, 19(1):33–46, 2007. <http://urisa.org/goldberg>.
34. Ching-Chien Chen, Craig A. Knoblock, and Cyrus Shahabi. Automatically conflating road vector data with orthoimagery. *Geoinformatica*, 10(4):495–530, December 2006.

35. Ion Muslea, Steve Minton, and Craig A. Knoblock. Active learning with multiple views. *Journal of Artificial Intelligence Research*, 27:203–233, 2006.
36. Shou de Lin and Craig A. Knoblock. Sergeant: A framework for building more flexible web agents by exploiting a search engine. *Web Intelligence and Agent Systems*, 3(1):1–15, 2005.
37. Greg Barish and Craig A. Knoblock. An expressive language and efficient execution system for software agents. *Journal of Artificial Intelligence Research*, 23:625–666, 2005.
38. Snehal Thakkar, Jose Luis Ambite, and Craig A. Knoblock. Composing, optimizing, and executing plans for bioinformatics web services. *VLDB Journal, Special Issue on Data Management, Analysis and Mining for Life Sciences*, 14(3):330–353, Sep 2005.
39. Kristina Lerman, Steven N. Minton, and Craig A. Knoblock. Wrapper maintenance: A machine learning approach. *Journal of Artificial Intelligence Research*, 18:149–181, 2003.
40. Naveen Ashish, Craig A. Knoblock, and Cyrus Shahabi. Selectively materializing data in mediators by analyzing user queries. *International Journal of Cooperative Information Systems*, 11(1), March, 2002.
41. Jose Luis Ambite and Craig A. Knoblock. Planning by rewriting. *Journal of Artificial Intelligence Research*, 15:207–261, 2001.
42. Sheila Tejada, Craig A. Knoblock, and Steven Minton. Learning object identification rules for information integration. *Information Systems*, 26(8), 2001.
43. Jose Luis Ambite, Craig A. Knoblock, Ion Muslea, and Andrew Philpot. Compiling source descriptions for efficient and flexible information integration. *Journal of Intelligent Information Systems*, 16(2):149–187, March 2001.
44. Craig A. Knoblock, Steven Minton, Jose Luis Ambite, Naveen Ashish, Ion Muslea, Andrew G. Philpot, and Sheila Tejada. The ariadne approach to web-based information integration. *International Journal of Cooperative Information Systems (IJCIS), Special Issue on Intelligent Information Agents: Theory and Applications*, 10(1/2):145–169, 2001.
45. Ion Muslea, Steven Minton, and Craig A. Knoblock. Hierarchical wrapper induction for semistructured information sources. *Autonomous Agents and Multi-Agent Systems*, 4(1/2), March 2001.
46. José Luis Ambite and Craig A. Knoblock. Flexible and scalable cost-based query planning in mediators: A transformational approach. *Artificial Intelligence Journal*, 118(1-2):115–161, April 2000.
47. Chun-Nan Hsu and Craig A. Knoblock. Semantic query optimization for query plans of heterogeneous multi-database systems. *IEEE Transactions on Knowledge and Data Engineering*, 12(6):959–978, November/December 2000.
48. Chun-Nan Hsu and Craig A. Knoblock. Discovering robust knowledge from databases that change. *Data Mining and Knowledge Discovery*, 2(1), 1997.
49. Yigal Arens, Craig A. Knoblock, and Wei-Min Shen. Query reformulation for dynamic information integration. *Journal of Intelligent Information Systems*, Special Issue on Intelligent Information Integration, 6(2/3):99–130, 1996.
50. Subbarao Kambhampati, Craig Knoblock, and Qiang Yang. Planning as refinement search: A unified framework for evaluating the design tradeoffs in partial order planning. *Artificial Intelligence*, 76(1-2), 1995.
51. Craig A. Knoblock. Automatically Generating Abstractions for Planning. *Artificial Intelligence*, 68(2), 1994.
52. Yigal Arens, Chin Y. Chee, Chun-Nan Hsu, and Craig A. Knoblock. Retrieving and integrating data from multiple information sources. *International Journal on Intelligent and Cooperative Information Systems*, 2(2):127–158, 1993.
53. Steven Minton, Jaime G. Carbonell, Craig A. Knoblock, Daniel R. Kuokka, Oren Etzioni, and Yolanda Gil. Explanation-based learning: A problem solving perspective. *Artificial Intelligence*, 40(1-3):63–118, 1989.

Books

1. Mayank Kejriwal, Craig A. Knoblock, and Pedro Szekely. Knowledge Graphs: Fundamentals, Techniques, and Applications. The MIT Press, Cambridge, MA, 2021.
2. Yao-Yi Chiang, Weiwei Duan, Stefan Leyk, Johannes Uhl, and Craig A. Knoblock. Using Historical Maps in Scientific Studies: Challenges and Best Practices. Springer International Publishing, 2020.
3. Craig A. Knoblock. Generating Abstraction Hierarchies. Kluwer Academic Publishers, Norwell, MA, 1993.

Book Chapters

1. Yao-Yi Chiang, Muhao Chen, Weiwei Duan, Jina Kim, Craig A Knoblock, Stefan Leyk, Zekun Li, Yijun Lin, Min Namgung, Basel Shbita, et al. GeoAI for the digitization of historical maps. In *Handbook of Geospatial Artificial Intelligence*, pages 217–247. CRC Press, 2023.
2. Rahul Parundekar, Craig A. Knoblock, and Jose Luis Ambite. Aligning ontologies of linked data. In *Linked Data Management: Principles and Techniques*. CRC Press, 2014.
3. Yao-Yi Chiang, Stefan Leyk, and Craig A. Knoblock. Efficient and robust graphics recognition from historical maps. In *Graphics Recognition: Achievements, Challenges, and Evolution, Selected Papers of the 8th International Workshop on Graphics Recognition (GREC), Lecture Notes in Computer Science*. Springer, 2012.
4. Kristina Lerman and Craig A. Knoblock. Wrapper maintenance. In Ling Liu and M. Tamer Ozsu, editors, *Encyclopedia of Database Systems*. Springer, Leipzig, Germany, 2009.
5. Yao-Yi Chiang and Craig A. Knoblock. Extracting road vector data from raster maps. In *Graphics Recognition: Achievements, Challenges, and Evolution, Selected Papers of the 8th International Workshop on Graphics Recognition (GREC), Lecture Notes in Computer Science, 6020*, pages 93–105. Springer, New York, 2009.
6. Daniel W. Goldberg, John P. Wilson, and Craig A. Knoblock. Extracting geographic features from the internet to automatically build detailed regional gazetteers. *International Journal of Geographic Information Science*, 23(1):92–128, 2009.
7. Ching-Chien Chen and Craig A. Knoblock. Conflation of geospatial data. In Shashi Shekhar and Hui Xiong, editors, *Encyclopedia of Geopatial Information Science*, pages 133–140. Springer, Leipzig, Germany, 2008. http://dx.doi.org/10.1007/978-0-387-35973-1_182.
8. Kristina Lerman and Craig A. Knoblock. Wrapper maintenance. In Ling Liu and M. Tamer Ozsu, editors, *Encyclopedia of Database Systems*. Springer, Leipzig, Germany, 2008.
9. Craig A. Knoblock and Cyrus Shahabi. Geospatial data integration. In John Wilson and A. Stewart Fotheringham, editors, *The Handbook of Geographic Information Science*. Wiley-Blackwell, 2007. ISBN: 978-1-4051-0796-9.
10. Mehdi Sharifzadeh, Cyrus Shahabi, and Craig A. Knoblock. Learning approximate thematic maps from labeled geospatial data. In Peggy Agouris and Arie Croitoru, editors, *Next Generation Geospatial Information: From Digital Image Analysis to SpatioTemporal Databases*, pages 129–141. A.A. Balkema Publishers, New York, 2005.
11. José Luis Ambite, Craig A. Knoblock, and Steve Minton. Plan optimization by plan rewriting. In Ioannis Vlahavas and Dimitris Vrakas, editors, *Intelligent Techniques for Planning*, pages 121–161. Idea Group Publishing, Hershey, PA, 2005.
12. Kristina Lerman, Steven N. Minton, and Craig A. Knoblock. Machine learning techniques for web wrapper maintenance. In Goran D. Putnik and Maria Manuela Cunha, editors, *Virtual Enterprise Integration: Technological and Organizational Perspectives*, pages 334–350. Idea Group Publishing, Hershey, PA, 2005.

13. Craig A. Knoblock, Kristina Lerman, Steven Minton, and Ion Muslea. Accurately and reliably extracting data from the web: A machine learning approach. In Piotr S. Szczepaniak, Javier Segovia, Janusz Kacprzyk, and Lotfi A. Zadeh, editors, *Intelligent Exploration of the Web*, pages 275–287. Springer-Verlag, Berkeley, CA, 2003.
14. Yigal Arens, Craig A. Knoblock, and Chun-Nan Hsu. Query processing in the sims information mediator. In Michael N. Huns and Munindar P. Singh, editors, *Readings in Agents*. Morgan Kaufmann, San Francisco, CA, 1998.
15. Craig A. Knoblock and Jose-Luis Ambite. Agents for information gathering. In J. Bradshaw, editor, *Software Agents*. AAAI/MIT Press, Menlo Park, CA, 1997.
16. Yigal Arens, Craig A. Knoblock, and Wei-Min Shen. Query reformulation for dynamic information integration. In Gio Wiederhold, editor, *Intelligent Integration of Information*. Kluwer Academic Publishers, Boston, MA, 1996.
17. Yigal Arens, Craig A. Knoblock, and Chun-Nan Hsu. Query processing in the SIMS information mediator. In Austin Tate, editor, *Advanced Planning Technology*. The AAAI Press, Menlo Park, CA, 1996.
18. Chun-Nan Hsu and Craig A. Knoblock. Using inductive learning to generate rules for semantic query optimization. In Gregory Piatetsky-Shapiro and Usama Fayyad, editors, *Advances in Knowledge Discovery and Data Mining*, chapter 17. MIT Press, Boston, MA, 1996.
19. Jaime Carbonell, Oren Etzioni, Yolanda Gil, Robert Joseph, Craig Knoblock, Steven Minton, and Manuela Veloso. Planning and learning in PRODIGY: Overview of an integrated architecture. In Ashwin Ram and David B. Leake, editors, *Goal-Driven Learning*, chapter 11, pages 297–306. MIT Press, Cambridge, MA, 1995.
20. Jaime G. Carbonell, Craig A. Knoblock, and Steven Minton. PRODIGY: An integrated architecture for planning and learning. In Kurt VanLehn, editor, *Architectures for Intelligence*. Erlbaum, Hillsdale, NJ, 1990.
21. Craig A. Knoblock. A theory of abstraction for hierarchical planning. In D. Paul Benjamin, editor, *Change of Representation and Inductive Bias*. Kluwer, Boston, MA, 1990.
22. Steven Minton, Jaime G. Carbonell, Craig A. Knoblock, Daniel R. Kuokka, Oren Etzioni, and Yolanda Gil. Explanation-based learning: A problem solving perspective. In Jaime G. Carbonell, editor, *Machine Learning: Paradigms and Methods*. MIT Press, Boston, MA, 1990.

Magazine Articles

1. Yolanda Gil, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankratius, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron Ridley, Hanan Samet, and Shashi Shekhar. Intelligent systems for geosciences: An essential research agenda. *Communications of the ACM*, 62(1):76–84, 2018.
2. Mayank Kejriwal, Pedro Szekely, and Craig A. Knoblock. Investigative knowledge discovery for combating illicit activities. *IEEE Intelligent Systems*, 33(1):53–63, 2018.
3. Craig A. Knoblock and Pedro Szekely. Exploiting semantics for big data integration. *AI Magazine*, 2015.
4. Craig A. Knoblock. Reduce data overload. *Earth Imaging Journal*, pages 28–30, March/April 2012.
5. Craig A. Knoblock, Jose Luis Ambite, Mark Carman, Matthew Michelson, Pedro Szekely, and Ratapoom Tuchinda. Beyond the elves: Making intelligent agents intelligent. *AI Magazine*, 29(2):33–42, 2008.

6. Martin Michalowski, Snehal Thakkar, and Craig A. Knoblock. Automatically utilizing secondary sources to align information across data sources. *AI Magazine, Special Issue on Semantic Integration*, 26(1), 2005.
7. Jose Luis Ambite, Craig A. Knoblock, Maria Muslea, and Steven Minton. Heracles ii: Conditional constraint networks for interleaved planning and information gathering. *IEEE Intelligent Systems*, 20(2):25–33, March/April 2005.
8. Martin Michalowski, Jose-Luis Ambite, Snehal Thakkar, Rattapoom Tuchinda, Craig A. Knoblock, and Steve Minton. Retrieving and semantically integrating heterogeneous data from the web. *IEEE Intelligent Systems*, 19(3), 2004.
9. Craig A. Knoblock. Agents for gathering, integrating, and monitoring information for travel planning. *IEEE Intelligent Systems*, 17(6), November/December 2002.
10. Hans Chalupsky, Yolanda Gil, Craig A. Knoblock, Kristina Lerman, Jean Oh, David V. Pynadath, Thomas A. Russ, and Milind Tambe. Electric elves: Agent technology for supporting human organizations. *AI Magazine*, 23(2):11–24, Summer 2002.
11. David N. Allsopp, Patrick Beautement, Michael Kirton, Jeffrey M. Bradshaw, Niranjan Suri, Edmund H. Durfee, Craig A. Knoblock, Austin Tate, and Craig W. Thompson. Coalition agents experiment: Multiagent cooperation in international coalitions. *IEEE Intelligent Systems*, 17(3):26–35, May/June 2002.
12. Craig A. Knoblock, Kristina Lerman, Steven Minton, and Ion Muslea. Accurately and reliably extracting data from the web: A machine learning approach. *IEEE Data Engineering Bulletin*, 23(4), December 2000.
13. Craig A. Knoblock and Steven Minton. The ariadne approach to web-based information integration. *IEEE Intelligent Systems*, 13(5), September/October 1998.
14. Jose Luis Ambite and Craig A. Knoblock. Agents for information gathering. *IEEE Expert: Intelligent Systems and their Applications*, September/October 1997.
15. Craig A. Knoblock and Qiang Yang. Relating the performance of partial-order planning algorithms to domain features. *SIGART Bulletin, Special Issue on Evaluating Plans, Planners and Planning Agents*, 6(1), 1995.

Conference Papers

1. Md Abrar Jahin, Craig A. Knoblock, and Jay Pujara. A Weisfeiler—Leman Characterization of Global-Attention Graph Transformers for Mixed-Integer Linear Programs. *Proceedings of the 2nd Conference on Topology, Algebra, and Geometry in Data Science (TAG-DS 2026)*, 334(2):123–142, 2026.
2. Md Abrar Jahin, Taufikur Rahman Fuad, Jay Pujara, and Craig A. Knoblock. ChronoSpike: An Adaptive Spiking Graph Neural Network for Dynamic Graphs. *Proceedings of the Fifth Learning on Graphs Conference (LOG 2026)*, 2026.
3. Craig A Knoblock, Binh Vu, Basel Shbita, Yao-Yi Chang, Xiao Lin, Goran Muric, Pothula Krishna, Jiyeon Pyo, Adriana Trejo-Sheu, and Meng Ye. Exploiting LLMs and semantic technologies to build a knowledge graph of historical mining data. In *The Semantic Web–ISWC 2025: 24th International Semantic Web Conference, ISWC 2025, November 2–6, 2025, Proceedings 20*. Springer International Publishing, 2025.
4. Binh Vu, Craig A Knoblock, and Fandel Lin. A domain-independent approach for semantic table interpretation. In *The Semantic Web–ISWC 2025: 24th International Semantic Web Conference, ISWC 2025, November 2–6, 2025, Proceedings 20*. Springer International Publishing, 2025.
5. Fandel Lin, Craig A. Knoblock, Binh Vu, and Yao-Yi Chiang. Exploiting polygon metadata to recolor historical maps. In *Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems, SIGSPATIAL ’25*, page 1122–1125, New York, NY, USA, 2025. Association for Computing Machinery.

6. Fandel Lin, Craig A. Knoblock, Binh Vu, Basel Shbita, and Yao-Yi Chiang. Exploiting polygon meta-data to colorize draft maps. In *Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '25, page 1126–1129, New York, NY, USA, 2025. Association for Computing Machinery.
7. Weiwei Duan, Yao-Yi Chiang, and Craig A. Knoblock. Ldtr: Linear object detection transformer for accurate graph generation by learning the n-hop connectivity information. In Xu-Cheng Yin, Dimosthenis Karatzas, and Daniel Lopresti, editors, *Document Analysis and Recognition – ICDAR 2025*, pages 40–59. Springer Nature Switzerland, 2025.
8. Weiwei Duan, Yao-Yi Chiang, Theresa Chen, Michael P. Gerlek, Leeje Jang, Sofia Kirsanova, Craig A. Knoblock, Fandel Lin, Yijun Lin, Zekun Li, and Steven N. Minton. Digmapper: A modular system for automated geologic map digitization. In *Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '25, page 717–728, New York, NY, USA, 2025. Association for Computing Machinery.
9. Fandel Lin, Craig A. Knoblock, and Binh Vu. A hierarchical voronoi approach to deploying new charging stations in an existing network. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '24, page 741–744, New York, NY, USA, 2024. Association for Computing Machinery.
10. Fandel Lin, Han Zhang, T. K. Satish Kumar, and Craig A. Knoblock. An integrated approach to multi-agent scheduling with bounded objectives. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '24, page 493–496, New York, NY, USA, 2024. Association for Computing Machinery.
11. Binh Vu, Craig A Knoblock, Basel Shbita, and Fandel Lin. Exploiting distant supervision to learn semantic descriptions of tables with overlapping data. In *The Semantic Web–ISWC 2024: 23th International Semantic Web Conference, ISWC 2024, November 11–15, 2024*. Springer International Publishing, 2024.
12. Basel Shbita and Craig A Knoblock. Automatically constructing geospatial feature taxonomies from openstreetmap data. In *2024 IEEE 18th International Conference on Semantic Computing (ICSC)*, pages 208–211. IEEE, 2024.
13. Minh Pham, Craig A. Knoblock, and Muhao Chen. Detecting semantic errors in tables using textual evidence. In *Proceedings of the 2023 IEEE International Conference on Big Data*, pages 292–303, 2023.
14. Fandel Lin and Craig A. Knoblock. Indirect cooperation in distributed stationary-resource searching with predefined destinations. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '23, New York, NY, USA, 2023. Association for Computing Machinery.
15. Fandel Lin, Craig A. Knoblock, Basel Shbita, Binh Vu, Zekun Li, and Yao-Yi Chiang. Exploiting polygon metadata to understand raster maps - accurate polygonal feature extraction. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '23, New York, NY, USA, 2023. Association for Computing Machinery.
16. Weiwei Duan, Yao-Yi Chiang, Stefan Leyk, Johannes H Uhl, and Craig A Knoblock. A label correction algorithm using prior information for automatic and accurate geospatial object recognition. In *Proceedings of the 2021 IEEE International Conference on Big Data*, pages 1604–1610. IEEE, 2021.
17. Weiwei Duan, Yao-Yi Chiang, Stefan Leyk, Johannes H Uhl, and Craig A Knoblock. Guided generative models using weak supervision for detecting object spatial arrangement in overhead images. In *Proceedings of the 2021 IEEE International Conference on Big Data*, pages 725–734. IEEE, 2021.
18. Minh Pham, Craig A. Knoblock, Muhao Chen, Binh Vu, and Jay Pujara. Spade: A semi-supervised probabilistic approach for detecting errors intables. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 3543–3551, August 2021.

19. Binh Vu, Craig A. Knoblock, Pedro Szekely, Minh Pham, and Jay Pujara. A graph-based approach for inferring semantic descriptions of Wikipedia tables. In Andreas Hotho, Eva Blomqvist, Stefan Dietze, Achille Fokoue, Ying Ding, Payam Barnaghi, Armin Haller, Mauro Dragoni, and Harith Alani, editors, *The Semantic Web – ISWC 2021*, pages 304–320. Springer International Publishing, 2021.
20. Basel Shbita, Craig A. Knoblock, Weiwei Duan, Yao-Yi Chiang, Johannes H Uhl, and Stefan Leyk. Building linked spatio-temporal data from vectorized historical maps. In *European Semantic Web Conference*, pages 409–426. Springer, 2020.
21. Zekun Li, Yao-Yi Chiang, Sasan Tavakkol, Basel Shbita, Johannes H. Uhl, Stefan Leyk, and Craig A. Knoblock. An automatic approach for generating rich, linked geo-metadata from historical map images. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3290–3298, 2020.
22. Daniel Garijo, Deborah Khider, Varun Ratnakar, Yolanda Gil, Ewa Deelman, Rafael Ferreira da Silva, Craig Knoblock, Yao-Yi Chiang, Minh Pham, Jay Pujara, Binh Vu, Dan Feldman, Rajiv Mayani, Kelly Cobourn, Chris Duffy, Armen Kemanian, Lele Shu, Vipin Kumar, Ankush Khandelwal, Kshitij Tayal, Scott Peckham, Maria Stoica, Anna Dabrowski, Daniel Hardesty-Lewis, and Suzanne Pierce. An intelligent interface for integrating climate, hydrology, agriculture, and socioeconomic models. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, IUI ’19, pages 111–112, New York, NY, USA, 2019. ACM.
23. Minh Pham, Craig A Knoblock, and Jay Pujara. Learning data transformations with minimal user effort. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 657–664. IEEE, 2019.
24. Binh Vu, Craig Knoblock, and Jay Pujara. Learning semantic models of data sources using probabilistic graphical models. In *The World Wide Web Conference, WWW ’19*, pages 1944–1953, New York, NY, USA, 2019. ACM.
25. Yuan Shi, TK Satish Kumar, and Craig A Knoblock. Automatic adaptation to sensor replacements. In *The Thirty-Second International Flairs Conference*, 2019.
26. Binh Vu, Jay Pujara, and Craig A Knoblock. D-repr: A language for describing and mapping diversely-structured data sources to rdf. In *Proceedings of the 10th International Conference on Knowledge Capture*, pages 189–196, 2019.
27. Basel Shbita, Arun Rajendran, Jay Pujara, and Craig A. Knoblock. Parsing, representing and transforming units of measure. In *Proceedings of the Conference on Modeling the World’s Systems*, 2019.
28. Yuan Shi, TK Satish Kumar, and Craig A Knoblock. Constraint-based learning for sensor failure detection and adaptation. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 328–335. IEEE, 2018.
29. Yolanda Gil, Kelly Cobourn, Ewa Deelman, Chris Duffy, Rafael Ferreira da Silva, Armen Kemanian, Craig Knoblock, Vipin Kumar, Scott Peckham, Lucas Carvalho, Yao-Yi Chiang, Daniel Garijo, Deborah Khider, Ankush Khandelwal, Minh Pahm, Jay Pujara, Varun Ratnakar, Maria Stoica, and Binh Vu. Mint: Model integration through knowledge-powered data and process composition. In *Proceedings of the 9th International Congress on Environmental Modelling and Software*, 2018.
30. T. K. Satish Kumar, Hong Xu, Zheng Tang, Anoop Kumar, Craig Milo Rogers, and Craig A. Knoblock. Alert generation in execution monitoring using resource envelopes. In *Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pages 38–43, 2018.
31. T. K. Satish Kumar, Zhi Wang, Anoop Kumar, Craig Rogers, and Craig A. Knoblock. Load scheduling of simple temporal networks under dynamic resource pricing. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-2018)*, 2018.
32. T. K. Satish Kumar, Hong Xu, Zheng Tang, Anoop Kumar, Craig Milo Rogers, and Craig A. Knoblock. A distributed logical filter for connected row convex constraints. In *Proceedings of the 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017.
33. Johannes Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan, and Craig A. Knoblock. Extracting human settlement footprint from historical topographic map series using context-based machine learning. In *8th International Conference on Pattern Recognition Systems, At Madrid, Spain, Volume: 8*, 2017.

34. Craig A. Knoblock, Pedro Szekely, Eleanor Fink, David Newbury Duane Degler, Robert Sanderson, Kate Blanch, Sara Snyder, Nilay Chheda, Nimesh Jain, Ravi Raju Krishna, Nikhila Begur Sreekanth, and Yixiang Yao. Lessons learned in building linked data for the american art collaborative. In *ISWC 2017 - 16th International Semantic Web Conference*, 2017.
35. Yuan Shi and Craig A. Knoblock. Learning with previously unseen features. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.
36. Linhong Zhu, Majid Ghasemi-Gol, Pedro Szekely, Aram Galstyan, and Craig A. Knoblock. Unsupervised entity resolution on multi-type graphs. In *ISWC 2016 - 15th International Semantic Web Conference*, 2016. Best Research Paper.
37. Ying Zhang, Yao-Yi Chiang, Craig A. Knoblock, Chaopeng Li, Liming Du, Shaowen Liu, and Sanjay Singh. An automatic approach for building place-name datasets from the web. In *Proceedings of the 19th AGILE International Conference on Geographic Information Science*, 2016.
38. Minh Pham, Suresh Alse, Craig Knoblock, and Pedro Szekely. Semantic labeling: A domain-independent approach. In *ISWC 2016 - 15th International Semantic Web Conference*, 2016.
39. Mohsen Taheriyani, Craig Knoblock, Pedro Szekely, and JosÈ Luis Ambite. Leveraging linked data to discover semantic relations within data sources. In *ISWC 2016 - 15th International Semantic Web Conference*, 2016.
40. Gleb Gawriljuk, Andreas Harth, Craig A. Knoblock, and Pedro Szekely. A scalable approach to incrementally building knowledge graphs. In *TPDL 2016 - 20th International Conference on Theory and Practice of Digital Libraries*, 2016.
41. Christian Paul, Achim Rettinger, Aditya Mogadala, Craig A. Knoblock, and Pedro Szekely. Efficient graph-based document similarity. In *The Semantic Web. Latest Advances and New Domains. 13th Extended Semantic Web Conference (ESWC), Crete, Greece.*, 2016.
42. Bo Wu and Craig A. Knoblock. Maximizing correctness with minimal user effort to learn data transformations. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 2016.
43. Pedro Szekely, Craig A. Knoblock, Jason Slepicka, Andrew Philpot, Amandeep Singh, Chengye Yin, Dipsy Kapoor, Prem Natarajan, Daniel Marcu, Kevin Knight, David Stallard, Subessware S. Karunamoorthy, Rajagopal Bojanapalli, Steven Minton, Brian Amanatullah, Todd Hughes, Mike Tamayo, David Flynt, Rachel Artiss, Shih-Fu Chang, Tao Chen, Gerald Hiebel, and Lidia Ferreira. Building and using a knowledge graph to combat human trafficking. In *Proceedings of the 14th International Semantic Web Conference (ISWC 2015)*, 2015.
44. Bo Wu and Craig A. Knoblock. An iterative approach to synthesize data transformation programs. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.
45. Craig A. Knoblock and Pedro Szekely. A scalable architecture for extracting, aligning, linking, and visualizing multi-int data. In *Proceedings of the Conference on the Next Generation Analyst III*. SPIE, 9499, 2015.
46. S.K. Ramnandan, Amol Mittal, Craig A. Knoblock, and Pedro Szekely. Assigning semantic labels to data sources. In *Proceedings of the 12th ESWC*, 2015.
47. Anastasia Dimou, Miel Vander Sande, Jason Slepicka, Pedro Szekely, Erik Mannens, Craig A. Knoblock, and Rik Van de Walle. Mapping hierarchical sources into RDF using the RML mapping language. In *IEEE International Conference on Semantic Computing (ICSC)*, pages 151–158. IEEE, 2014.
48. Eleanor E. Fink, Pedro Szekely, and Craig A. Knoblock. How linked open data can help in locating stolen or looted cultural property. In Marinos Ioannides, Nadia Magnenat-Thalmann, Eleanor Fink, Roko Zarnic, Alex-Yianing Yen, and Ewald Quak, editors, *Digital Heritage: Progress in Cultural Heritage: Documentation, Preservation, and Protection, LNCS 8740*, pages 228–237. Springer International Publishing, 2014.
49. Shrikanth Narayanan, Ayush Jaiswal, Yao-Yi Chiang, Yanhui Geng, Craig A. Knoblock, and Pedro Szekely. Integration and automation of data preparation and data mining. In *Proceedings of the First*

Workshop on Data Integration and Application at the 2014 IEEE International Conference on Data Mining, 2014.

50. Ayush Jaiswal, Yao-Yi Chiang, Craig A. Knoblock, and Liang Lan. Location prediction with sparse GPS data. In *Proceedings of the 8th Geographic Information Science*, 2014.
51. Mihir Sathe, Craig A. Knoblock, Yao-Yi Chiang, and Aaron Harris. A parallel query engine for interactive spatiotemporal analysis. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2014.
52. Mohsen Taheriyan, Craig A. Knoblock, Pedro Szekely, and Jose Luis Ambite. A Scalable Approach to Learn Semantic Models of Structured Sources. In *Proceedings of the 8th IEEE International Conference on Semantic Computing (ICSC 2014)*, 2014.
53. Bo Wu, Pedro Szekely, and Craig A. Knoblock. Minimizing user effort in transforming data by example. In *Proceedings of the International Conference on Intelligent User Interface*, 2014.
54. Rahul Parundekar, Craig A. Knoblock, and Jose Luis Ambite. Discovering alignments in ontologies of linked data. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2013.
55. Mohsen Taheriyan, Craig A. Knoblock, Pedro Szekely, and Jose Luis Ambite. A graph-based approach to learn semantic descriptions of data sources. In *Proceedings of the 12th International Semantic Web Conference (ISWC 2013)*, 2013.
56. Pedro Szekely, Craig A. Knoblock, Fengyu Yang, Xuming Zhu, Eleanor Fink, Rachel Allen, and Georgina Goodlander. Connecting the smithsonian american art museum to the linked data cloud. In *Proceedings of the 10th Extended Semantic Web Conference*, Montpellier, May 2013. Awarded Best In-Use Paper at ESWC 2013.
57. Aman Goel, Craig A. Knoblock, and Kristina Lerman. Exploiting Structure within Data for Accurate Labeling Using Conditional Random Fields. In *Proceedings of the 14th International Conference on Artificial Intelligence (ICAI)*, 2012.
58. Mohsen Taheriyan, Craig A. Knoblock, Pedro Szekely, and Jose Luis Ambite. Rapidly integrating services into the linked data cloud. In *Proceedings of the 11th International Semantic Web Conference (ISWC 2012)*, 2012.
59. Rahul Parundekar, Craig A. Knoblock, and Jose Luis Ambite. Discovering concept coverings in ontologies of linked data sources. In *Proceedings of the 11th International Semantic Web Conference (ISWC 2012)*, 2012. Awarded Best Research Paper at ISWC 2012.
60. Yao-Yi Chiang and Craig A. Knoblock. Generating named road vector data from raster maps. In *Proceedings of the Seventh International Conference on Geographic Information Science*, 2012.
61. Craig A. Knoblock, Pedro Szekely, Jose Luis Ambite, Shubham Gupta, Aman Goel, Maria Muslea, Kristina Lerman, and Parag Mallick. Semi-automatically mapping structured sources into the semantic web. In *Proceedings of the Extended Semantic Web Conference*, Crete, Greece, 2012.
62. Steven N. Minton, Sofus A. Macskassy, Peter LaMonica, KaneSee, Craig A. Knoblock, Greg Barish, Matthew Michelson, and Raymond Liuzzi. Monitoring entities in an uncertain world: Entity resolution and referential integrity. In *Proceedings of the 25th National Conference on Artificial Intelligence (AAAI-11)*, San Francisco, CA, 2011.
63. Yolanda Gil, Pedro Szekely, Sandra Villamizar, Thomas C. Harmon, Varun Ratnakar, Shubham Gupta, Maria Muslea, Fabio Silva, and Craig A. Knoblock. Mind your metadata: Exploiting semantics for configuration, adaptation, and provenance in scientific workflows. In *Proceedings of the 10th International Semantic Web Conference (ISWC 2011)*, 2011.
64. Yao-Yi Chiang and Craig A. Knoblock. Recognition of multi-oriented, multi-sized, and curved text. In *Proceedings of the Tenth International Conference on Document Analysis and Recognition*, 2011.
65. Yao-Yi Chiang and Craig A. Knoblock. A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition*, 2011.

66. Aman Goel, Craig A. Knoblock, and Kristina Lerman. Using conditional random fields to exploit token structure and labels for accurate semantic annotation. In *Proceedings of the 25th National Conference on Artificial Intelligence (AAAI-11)*, San Francisco, CA, 2011.
67. Shubham Gupta and Craig A. Knoblock. A framework for integrating and reasoning about geospatial data. In *Extended Abstracts of the Sixth International Conference on Geographic Information Science (GIScience)*, 2010.
68. Rahul Parundekar, Craig A. Knoblock, and José Luis Ambite. Linking and building ontologies of linked data. In *Proceedings of the 9th International Semantic Web Conference (ISWC 2010)*, 2010.
69. Yao-Yi Chiang and Craig A. Knoblock. An approach for recognizing text labels in raster maps. In *Proceedings of the 20th International Conference on Pattern Recognition*, 2010.
70. Shubham Gupta and Craig A. Knoblock. Building geospatial mashups to visualize information for crisis management. In *Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management*, 2010.
71. Craig A. Knoblock, Ching-Chien Chen, Yao-Yi Chiang, Aman Goel, Matthew Michelson, and Cyrus Shahabi. A general approach to discovering, registering, and extracting features from raster maps. In *Proceedings of the Conference on Document Recognition and Retrieval XVII of SPIE-IS&T Electronic Imaging*, volume 7534, 2010.
72. Yao-Yi Chiang and Craig A. Knoblock. Classification of raster maps for automatic feature extraction. In *GIS '09: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 138–147, New York, NY, USA, 2009. ACM.
73. Matthew Michelson and Craig A. Knoblock. Exploiting background knowledge to build reference sets for information extraction. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-2009)*, Pasadena, CA, 2009.
74. José Luis Ambite, Sirish Darbha, Aman Goel, Craig A. Knoblock, Kristina Lerman, Rahul Parundekar, and Thomas Russ. Automatically constructing semantic web services from online sources. In *Proceedings of the 8th International Semantic Web Conference (ISWC 2009)*, 2009.
75. Yao-Yi Chiang and Craig A. Knoblock. A method for automatically extracting road layers from raster maps. In *Proceedings of the Tenth International Conference on Document Analysis and Recognition*, 2009.
76. Matthew Michelson and Craig A. Knoblock. Mining the heterogeneous transformations between data sources to aid record linkage. In *Proceedings of the International Conference on Artificial Intelligence (ICAI)*, 2009.
77. Zachary G. Ives, Craig A. Knoblock, Steven Minton, Marie Jacob, Partha Pratim Talukdar, Rattapoom Tuchinda, José Luis Ambite, Maria Muslea, and B. Cenk Gazen. Interactive data integration through smart copy & paste. In *Fourth Biennial Conference on Innovative Data Systems Research (CIDR)*, Pacific Grove, CA, January 2009.
78. Matthew Michelson, Aman Goel, and Craig A. Knoblock. Identifying maps on the World Wide Web. In *Proceedings of the 5th International Conference on GIScience, LNCS 5266*, pages 249–260, New York, 2008. Springer. DOI=10.1007/978-3-540-87473-7_16.
79. Yao-Yi Chiang and Craig A. Knoblock. Automatic extraction of road intersection position, connectivity, and orientation from raster maps. In *16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS 2008)*, 2008.
80. Rattapoom Tuchinda, Pedro Szekely, and Craig A. Knoblock. Building mashups by example. In *Proceedings of the 2008 International Conference on Intelligent User Interface*, January 2008.
81. Martin Michalowski, Craig A. Knoblock, Kenneth M. Bayer, and Berthe Y. Choueiry. Exploiting Automatically Inferred Constraint Models for Building Identification in Satellite Imagery. In *Proceedings of the 15th ACM International Symposium on Advances in Geographic Information Systems (ACMGIS 07)*, pages 35–42, 2007.

82. Snehal Thakkar, Craig A. Knoblock, and Jose Luis Ambite. Quality-driven geospatial data integration. In *Proceedings of the 15th ACM International Symposium on Advances in Geographic Information Systems (ACMGIS 07)*, 2007.
83. Kenneth M. Bayer, Martin Michalowski, Berthe Y. Choueiry, and Craig A. Knoblock. Reformulating CSPs for Scalability with Application to Geospatial Reasoning. In *International Conference on Principles and Practice of Constraint Programming (CP 07)*, pages 164–179. LNCS 4741, Springer, 2007.
84. Rattapoom Tuchinda, Pedro Szekely, and Craig A. Knoblock. Building data integration queries by demonstration. In *Proceedings of the International Conference on Intelligent User Interface*, January 2007.
85. Mark James Carman and Craig A. Knoblock. Learning semantic descriptions of web information sources. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, January 2007.
86. Matthew Michelson and Craig A. Knoblock. Learning blocking schemes for record linkage. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, Boston, MA, 2006.
87. Kristina Lerman, Anon Plangrasopchok, and Craig A. Knoblock. Automatically labeling the inputs and outputs of web services. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, Boston, MA, 2006.
88. Yao-Yi Chiang and Craig A. Knoblock. Classification of line and character pixels on raster maps using discrete cosine transformation coefficients and support vector machines. In *Proceedings of the International Conference on Pattern Recognition (ICPR 2006)*, 2006.
89. Cyrus Shahabi, Yao-Yi Chiang, Kelvin Chung, Kai-Chen Huang, Jeff Khoshgozaran-Haghighi, Craig Knoblock, Sung Chun Lee, Ulrich Neumann, Ram Nevatia, Arjun Rihan, Snehal Thakkar, and Suyu You. Geodec: Enabling geospatial decision making. In *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME)*, Toronto, Canada, July 2006.
90. Craig A. Knoblock, Cyrus Shahabi, Ching-Chien Chen, and E. Lynn User. Automatic alignment of vector data and orthoimagery for the national map. In *Proceedings of the 7th Annual Conference on Digital Government Research*, San Diego, CA, 2006.
91. Craig A. Knoblock. Learning to optimize plan execution in information agents. In Hector Munoz-Avila and Francesco Ricci, editors, *Proceedings of the 6th International Conference on Case-Based Reasoning (ICCBR 2005)*, LNAI 3620, pages 2–3. Springer-Verlag, New York, 2005.
92. Yao-Yi Chiang, Craig A. Knoblock, and Ching-Chien Chen. Automatic extraction of road intersections from raster maps. In *The 13th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS'05)*, Bremen, Germany, November, 2005.
93. Martin Michalowski and Craig A. Knoblock. A constraint satisfaction approach to geospatial reasoning. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*, 2005.
94. Steven N. Minton, Claude Nanjo, Craig A. Knoblock, Martin Michalowski, and Matthew Michelson. Heterogeneous field matching method for record linkage. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM '05)*, 2005.
95. Matthew Michelson and Craig A. Knoblock. Semantic annotation of unstructured and ungrammatical text. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005)*, Edinburgh, Scotland, 2005.
96. Rattapoom Tuchinda and Craig A. Knoblock. Interactively building agents for consumer-side data mining. In *Proceedings of Intelligent User Interfaces*, San Diego, CA, 2005.
97. Ching-Chien Chen, Craig A. Knoblock, Cyrus Shahabi, Snehal Thakkar, and Yao-Yi Chiang. Automatically and accurately conflating orthoimagery and street maps. In *Proceedings of the 12th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS'04)*, 2004.

98. Rahul Bakshi, Craig A. Knoblock, and Snehal Thakkar. Exploiting online sources to accurately geocode addresses. In *Proceedings of the 12th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS'04)*, 2004.
99. Kristina Lerman, Lise Getoor, Steven Minton, and Craig A. Knoblock. Using the Structure of Web Sites for Automatic Segmentation of Tables. In *Proceedings of ACM SIG on Management of Data (SIGMOD-2004)*, 2004.
100. Craig A. Knoblock. Building software agents for planning, monitoring, and optimizing travel. In Andrew J. Frew, editor, *Proceedings of the Eleventh International Conference on Information Technology and Travel & Tourism*. Springer-Verlag, New York, 2004.
101. Rattapoom Tuchinda and Craig A. Knoblock. Agent wizard: Building information agents by answering questions. In *Proceedings of Intelligent User Interfaces*, Island of Madeira, Portugal, 2004.
102. Shou de Lin and Craig A. Knoblock. Exploiting a search engine to develop more flexible web agents. In *Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence (WI2003)*, pages 54–60, Halifax, Canada, 2003. IEEE Computer Society.
103. Shahram Ghandeharizadeh, Craig A. Knoblock, Christos Papadopoulos, Cyrus Shahabi, Esam Alwaigait, Jose Luis Ambite, Min Cai, Ching-Chien Chen, Parikshit Pol, Rolfe Schmidt, Saihong Song, Snehal Thakkar, and Runfang Zhou. Proteus: A system for dynamically composing and intelligently executing web services. In *Proceedings of the First International Conference on Web Services (ICWS)*, Las Vegas, NV, June 2003.
104. Ching-Chien Chen, Snehal Thakkar, Craig A. Knoblock, and Cyrus Shahabi. Automatically annotating and integrating spatial datasets. In *Proceedings of the Eighth International Symposium on Spatial and Temporal Databases (SSTD 2003)*, volume 2750 of *Lecture Notes in Computer Science*, Springer, Berlin, 2003.
105. Craig A. Knoblock. Deploying information agents on the web. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-2003)*, Acapulco, Mexico, 2003.
106. Oren Etzioni, Craig A. Knoblock, Rattapoom Tuchinda, and Alexander Yates. To buy or not to buy: Mining airline fare data to minimize ticket purchase price. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.
107. Ion Muslea, Steven Minton, and Craig A. Knoblock. Active learning with strong and weak views: A case study on wrapper induction. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-2003)*, Acapulco, Mexico, 2003.
108. Greg Barish and Craig A. Knoblock. Learning value predictors for the speculative execution of information gathering plans. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-2003)*, Acapulco, Mexico, 2003.
109. Jose Luis Ambite, Greg Barish, Craig A. Knoblock, Maria Muslea, Jean Oh, and Steven Minton. Getting from here to there: Interactive planning and agent execution for optimizing travel. In *Proceedings of the Fourteenth Conference on Innovative Applications of Artificial Intelligence (IAAI-2002)*, pages 862–869, Edmonton, Alberta, Canada, 2002.
110. Sheila Tejada, Craig A. Knoblock, and Steven Minton. Learning domain-independent string transformation weights for high accuracy object identification. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2002)*, Edmonton, Alberta, Canada, 2002.
111. D.N. Allsopp, P. Beautement, J.M. Bradshaw, E.H. Durfee, M. Kirton, C.A. Knoblock, N. Suri, A. Tate, and C.W. Thompson. Coalition agents experiment: Multi-agent co-operation in an international coalition setting. In *Proceedings of the Second International Conference on Knowledge Systems for Coalition Operations (KSCO-2002)*, Toulouse, France, April 2002.
112. Ion Muslea, Steven Minton, and Craig A. Knoblock. Active + semi-supervised learning = robust multi-view learning. In *Proceedings of the 19th International Conference on Machine Learning (ICML-2002)*, pages 435–442, Sydney, Australia, 2002.

113. Ion Muslea, Steven Minton, and Craig A. Knoblock. Adaptive view validation: A first step towards automatic view detection. In *Proceedings of the 19th International Conference on Machine Learning (ICML-2002)*, pages 443–450, Sydney, Australia, 2002.
114. Greg Barish and Craig A. Knoblock. Speculative execution for information gathering plans. In *Proceedings of the Sixth International Conference on Artificial Intelligence Planning and Scheduling*, Toulouse, France, 2002.
115. Cyrus Shahabi, Mohammada R. Kolahdouzan, Snehal Thakkar, Jose Luis Ambite, and Craig A. Knoblock. Efficiently querying moving objects with pre-defined paths in a distributed environment. In *Proceedings of the Ninth ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS)*, Atlanta, GA, November 2001.
116. Hans Chalupsky, Yolanda Gil, Craig A. Knoblock, Kristina Lerman, Jean Oh, David V. Pynadath, Thomas A. Russ, and Milind Tambe. Electric elves: Applying agent technology to support human organizations. In *Proceedings of the Conference on Innovative Applications of Artificial Intelligence*, 2001.
117. Craig A. Knoblock, Jose Luis Ambite, Steven Minton, Cyrus Shahabi, Mohammad Kolahdouzan, Maria Muslea, Jean Oh, and Snehal Thakkar. Integrating the world: The worldinfo assistant. In *Proceedings of the 2001 International Conference on Artificial Intelligence (IC-AI 2001)*, Las Vegas, NV, 2001.
118. Craig A. Knoblock, Steven Minton, Jose Luis Ambite, Maria Muslea, Jean Oh, and Martin Frank. Mixed-initiative, multi-source information assistants. In *Proceedings of the World Wide Web Conference*, Hong Kong, May 2001.
119. Martin Frank, Maria Muslea, Jean Oh, Steve Minton, and Craig Knoblock. An intelligent user interface for mixed-initiative multi-source travel. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, Santa Fe, New Mexico, January 2001.
120. Greg Barish, Daniel DiPasquo, Craig A. Knoblock, and Steven Minton. Dataflow plan execution for software agents. In *Proceedings of the Fourth International Conference on Autonomous Agents (Agents-2000)*, Barcelona, Spain, 2000.
121. Greg Barish, Daniel DiPasquo, Craig A. Knoblock, and Steven Minton. A dataflow approach to agent-based information management. In *Proceedings of the 2000 International Conference on Artificial Intelligence (IC-AI 2000)*, Las Vegas, NV, 2000.
122. José Luis Ambite, Craig A. Knoblock, and Steven Minton. Learning plan rewriting rules. In *Proceedings of the Fifth International Conference on Artificial Intelligence Planning and Scheduling Systems*, Breckenridge, CO, 2000.
123. Greg Barish, Craig A. Knoblock, Yi-Shin Chen, Steven Minton, Andrew Philpot, and Cyrus Shahabi. The theaterloc virtual application. In *Proceedings of Twelfth Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-2000)*, Austin, Texas, 2000.
124. Ion Muslea, Steven Minton, and Craig A. Knoblock. Selective sampling with redundant views. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000)*, 2000.
125. Naveen Ashish, Craig A. Knoblock, and Cyrus Shahabi. Selectively materializing data in mediators by analyzing user queries. In *Fourth International Conference on Cooperative Information Systems (CoopIS)*, Edinburgh, Scotland, September 1999.
126. Ion Muslea, Steve Minton, and Craig A. Knoblock. A hierarchical approach to wrapper induction. In *Proceedings of the 3rd International Conference on Autonomous Agents 1999*, Seattle, WA, 1999.
127. Craig A. Knoblock, Steven Minton, Jose Luis Ambite, Naveen Ashish, Pragnesh Jay Modi, Ion Muslea, Andrew G. Philpot, and Sheila Tejada. Modeling web sources for information integration. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, Madison, WI, 1998.
128. José Luis Ambite and Craig A. Knoblock. Flexible and Scalable Query Planning in Distributed and Heterogeneous Environments. In *Proceedings of the Fourth International Conference on Artificial Intelligence Planning Systems*, Pittsburgh, PA, 1998.

129. José Luis Ambite and Craig A. Knoblock. Planning by rewriting: Efficiently generating high-quality plans. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, Providence, RI, 1997.
130. Naveen Ashish and Craig A. Knoblock. Semi-automatic wrapper generation for Internet information sources. In *Proceedings of the Second IFCIS International Conference on Cooperative Information Systems*, Kiawah Island, SC, 1997.
131. Naveen Ashish, Craig A. Knoblock, and Alon Levy. Information gathering plans with sensing actions. In Sam Steel and Rachid Alami, editors, *Recent Advances in AI Planning: 4th European Conference on Planning, ECP'97*. Springer-Verlag, New York, 1997.
132. B. Cenk Gazen and Craig A. Knoblock. Combining the expressiveness of UCPOP with the efficiency of Graphplan. In Sam Steel and Rachid Alami, editors, *Recent Advances in AI Planning: 4th European Conference on Planning, ECP'97*. Springer-Verlag, New York, 1997.
133. Chun-Nan Hsu and Craig A. Knoblock. Discovering robust knowledge from dynamic closed-world data. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, Portland, OR, 1996.
134. Craig A. Knoblock. Building a planner for information gathering: A report from the trenches. In *Proceedings of the Third International Conference on Artificial Intelligence Planning Systems*, Edinburgh, Scotland, 1996.
135. Craig A. Knoblock. Planning, executing, sensing, and replanning for information gathering. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, Canada, 1995.
136. Chun-Nan Hsu and Craig A. Knoblock. Estimating the robustness of discovered knowledge. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, Montreal, Canada, 1995.
137. Yigal Arens and Craig A. Knoblock. Intelligent caching: Selecting, representing, and reusing data in an information server. In *Proceedings of the Third International Conference on Cooperative Information Systems*, Gaithersburg, MD, 1994.
138. Craig A. Knoblock. Generating parallel execution plans with a partial-order planner. In *Proceedings of the Second International Conference on Artificial Intelligence Planning Systems*, Chicago, IL, 1994.
139. Craig A. Knoblock and Qiang Yang. Evaluating the tradeoffs in partial-order planning algorithms. In *Proceedings of the Canadian Artificial Intelligence Conference*, 1994.
140. Chun-Nan Hsu and Craig A. Knoblock. Rule induction for semantic query optimization. In *Proceedings of the Eleventh International Conference on Machine Learning*, New Brunswick, NJ, 1994.
141. Craig Knoblock, Yigal Arens, and Chun-Nan Hsu. Cooperating agents for information retrieval. In *Proceedings of the Second International Conference on Cooperative Information Systems*, Toronto, Canada, 1994.
142. Chun-Nan Hsu and Craig A. Knoblock. Reformulating Query Plans For Multidatabase Systems. In *Proceedings of the Second International Conference on Information and Knowledge Management*, Washington, D.C., 1993.
143. Yigal Arens and Craig A. Knoblock. Planning and reformulating queries for semantically-modeled multidatabase systems. In *Proceedings of the First International Conference on Information and Knowledge Management*, Baltimore, MD, 1992.
144. Craig A. Knoblock. An Analysis of ABSTRIPS. In *Proceedings of the First International Conference on Artificial Intelligence Planning Systems*, College Park, MD, 1992.
145. Craig A. Knoblock. Search reduction in hierarchical problem solving. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, Anaheim, CA, 1991.
146. Craig A. Knoblock, Steven Minton, and Oren Etzioni. Integrating abstraction and explanation-based learning in PRODIGY. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, Anaheim, CA, 1991.

147. Craig A. Knoblock, Josh D. Tenenbergs, and Qiang Yang. Characterizing abstraction hierarchies for planning. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, Anaheim, CA, 1991.
148. Craig A. Knoblock. Learning abstraction hierarchies for problem solving. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, Boston, MA, 1990.

Workshop Papers, Symposium Papers, and Abstracts

1. Md Abrar Jahin, Craig A. Knoblock, and Jay Pujara. ArticleMiner: Ontology-guided knowledge graph construction from scientific publications. In *6th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment co-located with The 25th International Semantic Web Conference, ISWC 2026*, 2026.
2. Basel Shbita, Binh Vu, Fandel Lin, and Craig A. Knoblock. Embedding spatial and semantic contexts for geo-entity typing in smart city applications. In *Companion Proceedings of the ACM on Web Conference 2025, WWW '25*, page 1724–1732, New York, NY, USA, 2025. Association for Computing Machinery.
3. Binh Vu, C Knoblock, and Fandel Lin. Results of GRAMS+ at SemTab 2024. In *Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching co-located with the 23rd International Semantic Web Conference (ISWC 2024)*. SemTab, 2024.
4. Basel Shbita, Namrata Sharma, Binh Vu, Fandel Lin, and Craig A. Knoblock. Constructing a knowledge graph of historical mining data. In *6th International Workshop on Geospatial Linked Data (GeoLD) Co-located with the 21st Extended Semantic Web Conference (ESWC)*. CEUR, 2024.
5. J. H. Uhl, S. Leyk, W. Duan, Z. Li, B. Shbita, Y.-Y. Chiang, and C. A. Knoblock. Towards the large-scale extraction of historical land cover information from historical maps. *Abstracts of the ICA*, 3:297, 2021.
6. Johannes H. Uhl, Stefan Leyk, Yao-Yi Chiang, and Craig A. Knoblock. Towards the automated large-scale reconstruction of past road networks from historical maps. *Computers, Environment and Urban Systems*, 94:101794, 2022.
7. Binh Vu and Craig A Knoblock. Sand: A tool for creating semantic descriptions of tabular sources. In *The Semantic Web: ESWC 2022 Satellite Events: Hersonissos, Crete, Greece, Proceedings*, pages 63–67. Springer, 2022
8. Zekun Li, Runyu Guan, Qianmu Yu, Yao-Yi Chiang, and Craig A. Knoblock. Synthetic map generation to provide unlimited training data for historical map text detection. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, GEOAI '21*, page 17–26, New York, NY, USA, 2021. Association for Computing Machinery.
9. Deborah Khider, Yolanda Gil, Kelly M Cobourn, Ewa Deelman, Christopher Duffy, Rafael Ferreira da Silva, Armen Kemanian, Craig Knoblock, Vipin Kumar, Scott Dale Peckham, et al. Mint: An intelligent interface for understanding the impacts of climate change on hydrological, agricultural and economic systems. In *AGU Fall Meeting 2019*. AGU, 2019.
10. Rahul Gupta, Jay Pujara, Craig A Knoblock, Shushyam M Sharanappa, Bharat Pulavarti, Gerard Hoberg, and Gordon Phillips. Feature selection methods for understanding business competitor relationships. In *Proceedings of the Fourth International Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets*, ACM, 2018.
11. Weiwei Duan, Yao-Yi Chiang, Craig A. Knoblock, Stefan Leyk, and Johannes H. Uhl. Automatic generation of precisely delineated geographic features from georeferenced historical maps using deep learning. In *The 22nd International Research Symposium on Computer-based Cartography and GI-Science (Autocarto/UCGIS)*, 2018.
12. T. K. Satish Kumar, Zhi Wang, Anoop Kumar, Craig Rogers, and Craig Knoblock. On the linear programming duals of temporal reasoning problems. In *Proceedings of the Fifteenth International Symposium on Artificial Intelligence and Mathematics (ISAIM-2018)*, 2018.

13. Chun Lin, Hang Su, Craig A. Knoblock, Yao-Yi Chiang, Weiwei Duan, Stefan Leyk, and Johannes H. Uhl. Building linked data from historical maps. In *Proceedings of the ISWC 2018 Workshop on Enabling Open Semantic Science (SemSci 2018)*, 2018.
14. Craig A. Knoblock, Aparna R. Joshi, Abhishek Megotia, Minh Pham, and Chelsea Ursaner. Automatic spatio-temporal indexing to integrate and analyze the data of an organization. In *UrbanGIS 2017: International Workshop on Smart Cities and Urban Analytics*, 2017.
15. Weiwei Duan, Yao-Yi Chiang, Craig A. Knoblock, Vinil Jain, Dan Feldman, Johannes H. Uhl, and Stefan Leyk. Automatic alignment of geographic features in contemporary vector data and historical maps. In *1st ACM SIGSPATIAL Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery (GeoAI 2017)*, 2017.
16. Eric Heiden, Gerard Hoberg, Craig A. Knoblock, Palak Modi, Gordon Phillips, Gaurangi Raul, and Pedro Szekely. Web text-based network industry classifications: Preliminary results. In *Proceedings of DSMM: Data Science for Macro-Modeling with Financial and Economic Datasets*, 2017.
17. Mohsen Taheriyan, Craig Knoblock, Pedro Szekely, José Luis Ambite, and Yinyi Chen. Leveraging Linked Data to Infer Semantic Relations within Structured Sources. In *Proceedings of the 6th International Workshop on Consuming Linked Data (COLD 2015)*, 2015.
18. Jason Slepicka, Chengye Yin, Pedro Szekely, and Craig Knoblock. KR2RML: An Alternative Interpretation of R2RML for Heterogenous Sources. In *Proceedings of the 6th International Workshop on Consuming Linked Data (COLD 2015)*, 2015.
19. Bo Wu and Craig A. Knoblock. Iteratively learning conditional statements in transforming data by example. In *Proceedings of the First Workshop on Data Integration and Application at the 2014 IEEE International Conference on Data Mining*, pages 1105–1112. IEEE, 2014.
20. Shrikanth Narayanan, Ayush Jaiswal, Yao-Yi Chiang, Yanhui Geng, Craig A. Knoblock, and Pedro Szekely. Integration and automation of data preparation and data mining. In *Proceedings of the First Workshop on Data Integration and Application at the 2014 IEEE International Conference on Data Mining*, 2014.
21. Andreas Harth, Craig Knoblock, Steffen Stadtmüller, Rudi Studer, and Pedro Szekely. On-the-fly integration of static and dynamic sources. In *Proceedings of the Fourth International Workshop on Consuming Linked Data (COLD2013)*, 2013.
22. Ying Zhang, Yao-Yi Chang, Pedro Szekely, and Craig A. Knoblock. A semantic approach to retrieving, linking, and integrating heterogeneous geospatial data. In *Proceedings of the 2013 IJCAI Workshop on Semantic Cities*, 2013.
23. Craig A. Knoblock and Pedro Szekely. Semantics for big data integration and analysis. In *Proceedings of the AAAI Fall Symposium on Semantics for Big Data*, 2013.
24. Rahul Parundekar, Craig A. Knoblock, and Jose Luis Ambite. Finding concept coverings in aligning ontologies of linked data. In *Proceedings of the ESWC 2012 Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data*, 2012.
25. Bo Wu, Pedro Szekely, and Craig A. Knoblock. Learning data transformation rules through examples: Preliminary results. In *Ninth International Workshop on Information Integration on the Web (IIWeb 2012)*, 2012.
26. Mohsen Taheriyan, Craig A. Knoblock, Pedro Szekely, and Jose Luis Ambite. Semi-automatically modeling web apis to create linked apis. In *Proceedings of the ESWC 2012 Workshop on Linked APIs*, 2012.
27. Pedro Szekely, Craig A. Knoblock, Shubham Gupta, Mohsen Taheriyan, and Bo Wu. Exploiting semantics of web services for geospatial data fusion. In *Proceedings of the SIGSPATIAL International Workshop on Spatial Semantics and Ontologies (SSO 2011)*, Chicago, IL, 2011.
28. Rahul Parundekar, Jose Luis Ambite, and Craig A. Knoblock. Aligning unions of concepts in ontologies of geospatial linked data. In *Proceedings of the Terra Cognita 2011 Workshop in Conjunction with the 10th International Semantic Web Conference*, Bonn, Germany, 2011.

29. Craig A. Knoblock, Pedro Szekely, Jose Luis Ambite, Shubham Gupta, Aman Goel, Maria Muslea, Kristina Lerman, and Parag Mallick. Interactively mapping data sources into the semantic web. In *Proceedings of the First International Workshop on Linked Science 2011 in Conjunction with the 10th International Semantic Web Conference*, Bonn, Germany, 2011.
30. Yao-Yi Chiang, Stefan Leyk, and Craig A. Knoblock. Harvesting cartographic information from historical maps: Efficient user intervention strategy for robust graphics recognition. In *Proceedings of the Ninth IAPR International Workshop on Graphics RECOgnition (GREC'11)*, 2011.
31. Rahul Parundekar, Craig A. Knoblock, and José Luis Ambite. Aligning geospatial ontologies on the linked data web. In *Proceedings of the Workshop on Linked Spatiotemporal Data*, 2010.
32. Rahul Parundekar, Craig A. Knoblock, and José Luis Ambite. Linking the deep web to the linked data web. In *Proceedings of the AAAI Spring Symposium on Linked Data Meets Artificial Intelligence*, Stanford, CA, 2010.
33. Yao-Yi Chiang and Craig A. Knoblock. An approach to automatic road vectorization of raster maps. In *Proceedings of the 8th IAPR International Workshop on Graphics RECOgnition (GREC'09)*, 2009.
34. José Luis Ambite, Bora Cenk Gazen, Craig A. Knoblock, Kristina Lerman, and Thomas Russ. Discovering and learning semantic models of online sources for information integration. In *Proceedings of the IJCAI Workshop on Information Integration on the Web*, Pasadena, CA, 2009.
35. Jose Luis Ambite, Craig A. Knoblock, Kristina Lerman, Anon Plangprasopchok, Thomas Russ, Cenk Gazen, Steven Minton, and Mark Carman. Exploiting data semantics to discover, extract, and model web sources. In *Proceedings of the First International Workshop on Semantic Aspects in Data Mining (SADM'08)*, 2008.
36. Martin Michalowski, Craig A. Knoblock, and Berthe Y. Choueiry. Exploiting problem data to enrich models of constraint problems. In *Proceedings of the 6th International Workshop On Constraint Modelling and Reformulation (ModRef07)*, pages 1–15, 2007.
37. James Blythe, Dipsy Kapoor, Craig A. Knoblock, Kristina Lerman, and Steven Minton. Information integration for the masses. In *Proceedings of the 6th International Workshop on Information Integration on the Web (IIWeb)*, 2007.
38. Matthew Michelson and Craig A. Knoblock. Mining heterogeneous transformations for record linkage. In *Proceedings of the 6th International Workshop on Information Integration on the Web (IIWeb)*, pages 68–73, 2007.
39. Martin Michalowski, Craig A. Knoblock, and Berthe Y. Choueiry. Reformulating Constraint Models Using Input Data. In *Symposium on Abstraction, Reformulation and Approximation (SARA 07)*, pages 402–403, Whistler, Canada, 2007. LNAI 4612, Springer.
40. Kenneth M. Bayer, Martin Michalowski, Berthe Y. Choueiry, and Craig A. Knoblock. Reformulating Constraint Satisfaction Problems to Improve Scalability. In *Symposium on Abstraction, Reformulation and Approximation (SARA 07)*, pages 64–79, Whistler, Canada, 2007. LNAI 4612, Springer.
41. Matthew Michelson and Craig A. Knoblock. Beginning to understand unstructured, ungrammatical text: An information integration approach. In *Proceedings of the AAAI Spring Symposium on Machine Reading*, 2007.
42. Matthew Michelson and Craig A. Knoblock. An automatic approach to semantic annotation of unstructured, ungrammatical sources: A first look. In *Proceedings of the 1st IJCAI Workshop on Analytics for Noisy Unstructured Text Data*, 2007.
43. Ching-Chien Chen, Cyrus Shahabi, Craig A. Knoblock, and Mohammad R. Kolahdouzan. Automatically and efficiently matching road networks with spatial attributes in unknown geometry systems. In *The VLDB 2006 Workshop on Spatio-Temporal Database Management (STDBM)*, Seoul, Korea, September 2006.
44. Mohammad R. Kolahdouzan, Ching-Chien Chen, Cyrus Shahabi, and Craig A. Knoblock. Geomatchmaker: Automatic and efficient matching of vector data with spatial attributes in unknown geometry systems. In *Proceedings of the UCGIS 2005 Summer Assembly*, Jackson Hole, Wyoming, June 2005.

45. Mark James Carman and Craig A. Knoblock. Inducing source descriptions for automated web service composition. In *Proceedings of the AAAI 2005 Workshop on Exploring Planning and Scheduling for Web Services, Grid, and Autonomic Computing, Technical Report WS-05-03*. AAAI Press, 2005.
46. Craig A. Knoblock, Pedro Szekely, and Rattapoom Tuchinda. A mixed-initiative system for building mixed-initiative systems. In *Proceedings of the AAAI Fall Symposium on Mixed-Initiative Problem-Solving Assistants*, 2005.
47. Sneha Desai, Craig A. Knoblock, Yao-Yi Chiang, Kandarp Desai, and Ching-Chien Chen. Automatically identifying and georeferencing street maps on the web. In *Proceedings of the 2nd International Workshop on Geographic Information Retrieval (GIR'05)*, 2005.
48. Ching-Chien Chen, Cyrus Shahabi, and Craig A. Knoblock. Utilizing road network data for automatic identification of road intersections from high resolution color orthoimagery. In *Proceedings of the 2nd Workshop on Spatio-Temporal Database Management - STDBM'04*, 2004.
49. Kristina Lerman, Cenk Gazen, Steven Minton, and Craig A. Knoblock. Populating the semantic web. In *Proceedings of the AAAI 2004 Workshop on Advances in Text Extraction and Mining*, 2004.
50. Martin Michalowski, Snehal Thakkar, and Craig A. Knoblock. Exploiting secondary sources for unsupervised record linkage. In *Proceedings of the 2004 VLDB Workshop on Information Integration on the Web*, 2004.
51. Snehal Thakkar, Jose Luis Ambite, and Craig A. Knoblock. A data integration approach to automatically composing and optimizing web services. In *Proceedings of 2004 ICAPS Workshop on Planning and Scheduling for Web and Grid Services*, Whistler, BC, Canada, 2004.
52. Ching-Chien Chen, Craig A. Knoblock, Cyrus Shahabi, and Snehal Thakkar. Automatically and accurately conflating satellite imagery and maps. In *Proceedings of the International Workshop on Next Generation Geospatial Information*, Cambridge, MA, 2003.
53. Mehdi Sharifzadeh, Cyrus Shahabi, and Craig A. Knoblock. Learning approximate thematic maps from labeled geospatial data. In *Proceedings of the International Workshop on Next Generation Geospatial Information*, Cambridge, MA, 2003.
54. Martin Michalowski, Snehal Thakkar, and Craig A. Knoblock. Exploiting secondary sources for automatic object consolidation. In *Proceedings of the KDD'03 Workshop on Data Cleaning, Record Linkage and Object Consolidation*, 2003.
55. Greg Barish and Craig A. Knoblock. Combining classification and transduction for value prediction in speculative plan execution. In *Proceedings of 2003 IJCAI Workshop on Information Integration on the Web*, Acapulco, Mexico, 2003.
56. Snehal Thakkar and Craig A. Knoblock. Efficient execution of recursive integration plans. In *Proceedings of 2003 IJCAI Workshop on Information Integration on the Web*, Acapulco, Mexico, 2003.
57. Snehal Thakkar, Jose-Luis Ambite, and Craig A. Knoblock. A view integration approach to dynamic composition of web services. In *Proceedings of 2003 ICAPS Workshop on Planning for Web Services*, Trento, Italy, 2003.
58. Greg Barish and Craig A. Knoblock. An efficient and expressive language for information gathering on the web. In *Proceedings of the AIPS-2002 Workshop on Is there life after operator sequencing? - Exploring real world planning*, pages 5–12, Toulouse, France, 2002.
59. Greg Barish and Craig A. Knoblock. Learning efficient value predictors for speculative plan execution. In *Proceedings of the Fifth International Workshop on the World Wide Web and Databases (WebDB 2002)*, pages 77–82, Madison, WI, 2002.
60. Ion Muslea, Steven Minton, and Craig A. Knoblock. Adaptive view validation: A case study on wrapper induction and text classification. In *Proceedings of the AAAI-2002 Workshop on Intelligent Services Integration*, Edmonton, Alberta, Canada, 2002.
61. Snehal Thakkar, Craig A. Knoblock, Jose Luis Ambite, and Cyrus Shahabi. Dynamically composing web services from on-line sources. In *Proceeding of AAAI-2002 Workshop on Intelligent Service Integration*, pages 1–7, Edmonton, Alberta, Canada, July 2002.

62. Kristina Lerman, Craig A. Knoblock, and Steven Minton. Automatic data extraction from lists and tables in web sources. In *Proceedings of the IJCAI 2001 Workshop on Adaptive Text Extraction and Mining*, Seattle, WA, 2001.
63. Greg Barish, Craig A. Knoblock, Yi-Shin Chen, Steven Minton, Andrew Philpot, and Cyrus Shahabi. Theaterloc: A case study in building an information integration application. In *Proceedings of the IJCAI-99 Workshop on Intelligent Information Integration*, Stockholm, Sweden, 1999.
64. Greg Barish, Craig A. Knoblock, Daniel DiPasquo, and Steven Minton. An efficient plan execution system for information management agents. In *Proceedings of the ACM CIKM'99 Workshop on Web Information and Data Management (WIDM)*, Kansas City, MO, 1999.
65. Craig A. Knoblock and Steven Minton. Building agents for internet-base supply chain integration. In *Proceedings of the Workshop on Agents for Electronic Commerce and Managing the Internet-Enabled Supply Chain*, Seattle, WA, 1999.
66. Ion Muslea, Steven Minton, and Craig A. Knoblock. Selective sampling with co-testing. In *Proceedings of the CRM Workshop on Selecting and Combining Models with Machine Learning Algorithms*, Montreal, Canada, April 2000.
67. Ion Muslea, Steven Minton, and Craig A. Knoblock. Selective sampling with naive co-testing: Preliminary results. In *Proceedings of the ECAI-2000 Workshop On Machine Learning for Information Extraction*, Berlin, Germany, 2000.
68. Craig A. Knoblock Naveen Ashish and Cyrus Shahabi. Selectively materializing data in mediators by analyzing query distribution, source structure and maintenance cost. In *Proceedings of the ACM CIKM'99 Workshop on Web Information and Data Management (WIDM)*, Kansas City, MO, 1999.
69. Naveen Ashish and Craig A. Knoblock. Wrapper generation for semi-structured Internet sources. In *Proceedings of the Workshop on Management of Semistructured Data*, Tucson, AZ, 1997.
70. Craig A. Knoblock. Why plan generation and plan execution are inseparable. In *Proceedings of the AAAI Fall Symposium on Plan Execution*, Cambridge, MA, 1996.
71. Sheila Tejada, Craig A. Knoblock, and Steven Minton. Learning models for multi-source integration. In *Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access*, Stanford, CA, 1996.
72. Craig A. Knoblock. Integrating planning and execution for information gathering. In *Working Notes of the AAAI Spring Symposium on Information Gathering in Distributed Heterogeneous Environments*, Palo Alto, CA, 1995.
73. Craig A. Knoblock and Alon Levy. Exploiting run-time information for efficient processing of queries. In *Working Notes of the AAAI Spring Symposium on Information Gathering in Distributed Heterogeneous Environments*, Palo Alto, CA, 1995.
74. Jose-Luis Ambite and Craig A. Knoblock. Reconciling distributed information sources. In *Working Notes of the AAAI Spring Symposium on Information Gathering in Distributed Heterogeneous Environments*, Palo Alto, CA, 1995.
75. Craig A. Knoblock and Alon Levy. Efficient query processing for information gathering agents. In *Proceedings of the Workshop on Intelligent Information Agents*, Gaithersburg, MD, 1994.
76. Jose Luis Ambite and Craig A. Knoblock. Reconciling agent models. In *Proceedings of the Workshop on Intelligent Information Agents*, Gaithersburg, MD, 1994.
77. Craig A. Knoblock. Applying a general-purpose planner to the problem of query access planning. In *Proceedings of the AAAI Fall Symposium on Planning and Learning: On to Real Applications*, New Orleans, La, 1994.
78. Craig A. Knoblock and Qiang Yang. Evaluating the tradeoffs in partial-order planning algorithms. In *Proceedings of the Workshop on Comparative Analysis of AI Planning Systems*, pages 42–58, 1994.
79. Craig A. Knoblock and Qiang Yang. A Comparison of the SNLP and TWEAK Planning Algorithms. In *Proceedings of the AAAI Spring Symposium on Classical Planning Systems*, Palo Alto, CA, 1993.

80. Chun-Nan Hsu and Craig A. Knoblock. Learning Database Abstractions for Query Reformulation. In *Proceedings of the AAAI Workshop on Knowledge Discovery in Databases*, Washington, DC, 1993.
81. Craig A. Knoblock, Josh D. Tenenbergh, and Qiang Yang. A spectrum of abstraction hierarchies for planning. In *Proceedings of the Workshop on Automatic Generation of Approximations and Abstractions*, Boston, MA, 1990.
82. Craig A. Knoblock. Abstracting the Tower of Hanoi. In *Proceedings of the Workshop on Automatic Generation of Approximations and Abstractions*, Boston, MA, 1990.
83. Craig A. Knoblock. Learning problem-specific abstraction hierarchies. In *Proceedings of the Workshop on Change of Representation and Problem Reformulation*, Palo Alto, CA, 1990.
84. Jaime G. Carbonell, Yolanda Gil, Robert Joseph, Craig A. Knoblock, Steve Minton, and Manuela M. Veloso. Designing an integrated architecture: The PRODIGY view. In *Proceedings of The Twelfth Annual Conference of the Cognitive Science Society*, Boston, MA, 1990.
85. Craig A. Knoblock. Learning hierarchies of abstraction spaces. In *Proceedings of the Sixth International Workshop on Machine Learning*, Ithaca, NY, 1989.
86. Craig A. Knoblock. Automatically generating abstractions for planning. In *Proceedings of the First International Workshop on Change of Representation and Inductive Bias*, Briarcliff, NY, 1988.
87. Steven Minton, Jaime G. Carbonell, Oren Etzioni, Craig A. Knoblock, and Daniel R. Kuokka. Acquiring effective search control rules: Explanation-based learning in the PRODIGY system. In *Proceedings of the Fourth International Workshop on Machine Learning*, Irvine, CA, 1987.
88. Steven Minton, Jaime G. Carbonell, Craig A. Knoblock, Dan Kuokka, and Henrik Nordin. Improving the effectiveness of explanation-based learning in the PRODIGY system. In *Proceedings of the Workshop on Knowledge Compilation*, Inn at Otter Crest, OR, 1986.

Demonstration Papers

1. Craig A Knoblock, Binh Vu, Basel Shbita, Pothula Punith Krishna, and Namrata Sharma. Demonstrating MinMod: A large-scale knowledge graph of historical mining data. In *International Semantic Web Conference*, 2025.
2. Jason Slepicka, Chengye Yin, Pedro Szekely, and Craig A. Knoblock. A demonstration of linked data source discovery and integration. In *Proceedings of the ISWC 2014 Posters & Demonstrations Track*, volume 1272. CEUR, 2014.
3. Yao-Yi Chiang, Sima Moghaddam, Sanjauli Gupta, Renuka Fernandes, and Craig A. Knoblock. From map images to geographic names. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2014.
4. Yao-Yi Chiang, Bo Wu, Akshay Anand, Ketan Akade, and Craig A. Knoblock. A system for efficient cleaning and transformation of geospatial data attributes. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2014.
5. Craig A. Knoblock, Pedro Szekely, Shubham Gupta, Animesh Manglik, Ruben Verborgh, Fengyu Yang, and Rik Van de Walle. Publishing data from the Smithsonian American Art Museum as linked open data. In *Proceedings of the ISWC 2013 Posters & Demonstrations Track*, pages 129–132, 2013.

Selected Technical Reports

1. Jose-Luis Ambite, Yigal Arens, Naveen Ashish, Chin Y. Chee, Chun-Nan Hsu, Craig A. Knoblock, Wei-Min Shen, and Sheila Tejada. The SIMS manual: Version 1.0. Technical Report ISI/TM-95-428, University of Southern California, Information Sciences Institute, 1995.
2. Craig A. Knoblock. *Automatically Generating Abstractions for Problem Solving*. Ph.D. Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1991.

3. Steven Minton, Craig A. Knoblock, Daniel R. Kuokka, Yolanda Gil, Robert L. Joseph, and Jaime G. Carbonell. *PRODIGY 2.0: The manual and tutorial*. Technical Report CMU-CS-89-146, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1989.

Tutorials

1. *Knowledge Graph Construction from Web Corpora*
Mayank Kejriwal, Craig Knoblock, and Pedro Szekely
Tutorial Program, 32nd AAAI Conference on Artificial Intelligence
February 3, 2018
2. *Constructing Domain-specific Knowledge Graphs*
Mayank Kejriwal, Pedro Szekely and Craig Knoblock
Tutorial Program, 16th International Semantic Web Conference
October 21, 2017
3. *Information Integration on the Web*
Craig Knoblock and Subbarao Kambhampati
Tutorial Program, Twenty-Second National Conference on Artificial Intelligence
July 22, 2007
4. *Planning on the Web*
Craig Knoblock and Jose Luis Ambite
Tutorial Program, Fourteenth International Conference on Automated Planning and Scheduling
June 3, 2004
5. *Information Integration on the Web*
Craig Knoblock and Subbarao Kambhampati
Tutorial Program, Eighteenth National Conference on Artificial Intelligence
July 29, 2002
6. *Planning and Plan Execution for Information Gathering*
Tutorial Program, PLANET International Summer School on AI Planning
September 16-22, 2002 Halkidiki, Greece
7. *Recent Advances in AI Planning*
Co-taught with Subbarao Kambhampati
Tutorial Program, Fifteenth National Conference on Artificial Intelligence, 1998.
8. *Information Gathering and Integration*
Co-taught with Alon Halevy
Tutorial Program, Thirteenth National Conference on Artificial Intelligence, 1996.

Grants, Contracts, & Gifts

1. **Principal Investigator**, Enhancing Scalability of Knowledge Graphs for Enterprise Defense Solutions, United States Air Force (subcontract via Simba Chain, Inc.), \$80,000, April 2026.
2. **Principal Investigator**, Modeling Sparse and Heterogenous Geochemistry Data to Accelerate Critical Mineral Discovery, DARPA (subcontract via Inferlink, Inc.), \$1,799,995, January 2026.
3. **Principal Investigator**, Extracting Models of Minerals from Knowledge, DARPA & USGS, \$1,000,000, August 2023.
4. **Principal Investigator**, AI for Map Geolocation and Extraction to Find Critical Minerals, DARPA & USGS (subcontract via Inferlink, Inc.), \$469,346, August 2023.
5. **Co-Principal Investigator**, Combining Efficient Algorithms, Machine Learning and Knowledge Graphs for Scalable, High-Dimensional Nearest Neighbor Search, NGA, \$250,000, March 2022.

6. **Principal Investigator**, Food & Agricultural Assurance & Supply Chains Testbed (FAAST), DARPA (subcontract via UMD ARLIS), \$400,000, April 2021.
7. **Co-Principal Investigator**, FAIR Data Platform for Scaling up Data For FAIR-fication, Novartis, \$613,208, August 2020.
8. **Co-Principal Investigator**, *Leveraging Financial and Economic Data: Business OKN*, National Science Foundation, \$999,117, September 2019.
9. **Principal Investigator**, *Computer Vision Artificial Intelligence Search Tool (CFAST)*, Combating Terrorism Technical Support Office (CTTSO) (subcontract via SRI), \$320,000, July 2019.
10. **Principal Investigator**, *Multi-Source Data Fusion for Space Situational Awareness (SpaceAware)*, Air Force, \$929,441, April 2018.
11. **Co-Principal Investigator**, *MINT: Model INTeGration through Knowledge-Rich Data and Process Composition*, DARPA, \$12,979,881, December 2017.
12. **Principal Investigator**, *Exploiting Context in Cartographic Evolutionary Documents to Extract and Build Linked Spatial-Temporal Datasets*, National Science Foundation, \$642,959, August 2016.
13. **Principal Investigator**, *Automatic Alignment of Design Semantics to Enable Mapping Between CAD Systems*, DARPA, \$120,000, September 2016.
14. **Co-Principal Investigator**, *Effectively Forecasting Evolving Cyber Threats*, IARPA, \$6,476,872, July 2016.
15. **Principal Investigator**, *American Art Collaborative's (AAC) Linked Open Data (LOD) Initiative*, Mellon Foundation (subcontract via Smithsonian American Art Museum), \$100,000, July 2016.
16. **Principal Investigator**, *Large-scale Database Construction of Firms, Organizational Form, Competition, and Industry Change*, National Science Foundation, \$250,000, March 2016.
17. **Principal Investigator**, *Phase 1 of the American Art Collaborative Linked Open Data Initiative*, Institute of Museum and Library Services (subcontract via Crystal Bridges Museum of American Art), \$50,000, January 2016.
18. **Principal Investigator**, *Probabilistic Representation of Intent Commitments to Ensure Software Survival*, DARPA (subcontract via Charles River Associates), \$1,294,800, November 2015.
19. **Principal Investigator**, *Organizing Data for Effective End-User Management*, U.S. Government, \$225,000, August 2015.
20. **Principal Investigator**, *Scalable Coordinated Human-centered Automated Resilient Planning*, DARPA (subcontract via Next Century), \$3,425,000, August 2015.
21. **Co-Principal Investigator**, *Domain-Specific Insight Graphs*, DARPA, \$8,400,000, September 2014.
22. **Co-Principal Investigator**, *Text and Symbol Recognition from Historical Maps, Conveyancing Liability Solutions*, \$120,000, September 2014.
23. **Principal Investigator**, *Creating Linked Open Data About Conflict Indicators for Peacebuilding*, Center for Global Health and Peacebuilding, \$100,000, July 2014.
24. **Principal Investigator**, *Extracting and Integrating User Generated Content*, National Geospatial-Intelligence Agency (subcontract via SRI), \$100,000, October 2014.
25. **Principal Investigator**, *Integrating and Visualizing Diverse Sources of Network Data, Phase II*, Air Force Research Lab (via Northrup Grumman), \$35,000, April 2014.
26. **Principal Investigator**, *Building Integration Tools for Complex Domains*, IARPA (via CUBRC), \$75,000, January 2014.
27. **Principal Investigator**, *Integrating Heterogeneous Sources in a Geospatial Framework to Support Oil Field Operations*, Chevron, \$116,000, January 2014.
28. **Principal Investigator**, *Semantic Alignment to Integration Data*, US Government, \$340,000, November 2013.

29. **Principal Investigator**, *Integrating and Visualizing Diverse Sources of Network Data*, Air Force Research Lab (via Northrup Grumman), \$125,000, October 2013.
30. **Principal Investigator**, *A Unified Approach to Information Integration and Data Mining on Large, Heterogeneous Data Sources*, Huawei Technologies, \$130,000, September 2013.
31. **Principal Investigator**, *Harvesting Geographic Information from Heterogeneous Raster Maps*, TerraGo Technologies, \$75,000, August 2013.
32. **Principal Investigator**, *Integrating Heterogeneous Sources in a Geospatial Framework to Support Oil Field Operations*, Chevron, \$110,000, January 2013.
33. **Principal Investigator**, *U.S. Based Students Travel Support for the ACM SIGSPATIAL 2013*, NSF, \$30,000, August 2013.
34. **Principal Investigator**, *Organizing and Linking Art Museum Information*, Smithsonian American Art Museum, \$200,000, April 2012.
35. **Principal Investigator**, *Integrating Heterogeneous Sources in a Geospatial Framework to Support Oil Field Operations*, Chevron, \$145,000, January 2012.
36. **Co-Principal Investigator**, *Geospatial Analysis of Motion-Based Intelligence Tracking*, Office of Naval Research, \$2,520,000, December 2011.
37. **Principal Investigator**, *Building Linked Open Services from Online Sources*, NSF, \$516,000, October 2011.
38. **Principal Investigator**, *Integrating Heterogeneous Sources in a Geospatial Framework to Support Oil Field Operations*, Chevron, \$96,000, January 2011.
39. **Principal Investigator**, *D0main Vocabulary from Extraction, Transduction, and Auto Induction of Layout (DOVETAIL)*, IARPA, (subcontract from SRI), \$2,307,822, October 2010.
40. **Principal Investigator**, *An Integrated Framework for Geospatial Data Fusion*, DOD, \$300K, August 2010.
41. **Principal Investigator**, **Building Mashups for Decision Making**, DARPA, \$150,000, April 2010.
42. **Principal Investigator**, *Integrating Heterogeneous Sources in a Geospatial Framework to Support Oil Field Operations*, Chevron, \$150,000, January 2010.
43. **Principal Investigator**, *Harvesting Named Geographic Features from Raster Maps*, DOD, \$300,000, October 2009.
44. **Co-Principal Investigator**, *Automation of Technical Reachback for PMESII Best Practices*, OSD STTR Phase I Project (subcontract from Securborator), \$100,000, February 2009.
45. **Co-Principal Investigator**, *Trend Analysis and Data Mining for Finding Suspicious Entities*, AFRL STTR Phase II Project (subcontract from Fetch Technologies), \$750,000, November 2008.
46. **Co-Principal Investigator**, *Rapid engagement of humanitarian assistance social networks for ad hoc geospatial data sharing*, National Science Foundation, \$450,000, September 2008.
47. **Principal Investigator**, *Sapling Effort on Information Integration*, DARPA, \$500,000, November 2007.
48. **Co-Principal Investigator**, *Exploiting Raster Maps for Imagery Analysis*, AFOSR STTR Phase II Project (subcontract from Geosemble Technologies), \$742,000, October 2007.
49. **Co-Principal Investigator**, *Trend Analysis and Data Mining for Finding Suspicious Entities*, AFRL STTR Phase I Project (subcontract from Fetch Technologies), \$100,000, October 2007.
50. **Principal Investigator**, *Geospatial Information Integration and Visualization for Oil Field Operation and Exploration*, Chevron, \$275,000, September 2007.
51. **Principal Investigator**, *Interleaved Information Gathering and Reasoning for Information Fusion*, AFOSR, \$740,000, April 2007.

52. **Principal Investigator**, *Integrating Online Maps with Aerial Imagery*, Microsoft Research, \$40,000, March 2007.
53. **Co-Principal Investigator**, *Exploiting Raster Maps for Imagery Analysis*, AFOSR STTR Phase I Project (subcontract from Geosemble Technologies), \$100,000, September 2006.
54. **Co-Principal Investigator**, *GeoDec: Geospatial Decision Making*, Microsoft Research, \$40,000, January 2006.
55. **Co-Principal Investigator**, *Accurate and Rapid Construction of Information-Rich 3-D City Models*, Microsoft Research, \$100,000, October 2005.
56. **Principal Investigator**, *PSYOP Global Reach*, Institute for Creative Technologies \$200,000, October 2005.
57. **Principal Investigator**, *EntityBases: Large-scale Knowledge Bases for Intelligence Data*, AFRL, \$3,450,000, April 2005.
58. **Principal Investigator**, *Constraint-based Integration of Geospatial and Online Sources*, Institute for Creative Technologies, \$501,000, March 2004.
59. **Principal Investigator**, *A Framework for Integrating Geospatial and Online Data to Respond to Unexpected Events*, NSF ITR Grant, \$1,180,000, October 2003.
60. **Principal Investigator**, *Enduring Personal Cognitive Assistant*, DARPA PAL Program (subcontract from SRI), \$3,125,000, May 2003.
61. **Co-Principal Investigator**, *Proteus: A System for Execution of Dynamically Composed Web Services Using GXA*, Microsoft, \$350,000, January 2003.
62. **Co-Principal Investigator**, *Machine Learning for Record Linkage II*, AFOSR STTR Phase II (subcontract from Fetch Technologies), \$173,000, September 2002.
63. **Principal Investigator**, *Efficiently Accessing and Integrating Worldwide Geographic Data (Tera-World)*, AFOSR DURIP equipment grant, \$208,000, April 2002.
64. **Co-Principal Investigator**, *Composing Web Sources Using a Scalable .Net Middleware*, Microsoft, \$200,000, January 2002.
65. **Co-Principal Investigator**, *Machine Learning for Record Linkage (ML-Link)*, AFOSR STTR Phase I project (subcontract from Fetch Technologies), \$36,000, October 2001.
66. **Principal Investigator**, *WIDELink: A Bootstrapping Approach to Identifying, Modeling, and Linking On-Line Data Sources*, DARPA Evidence Extraction and Link Discovery Program, \$1,485,000, September 2001.
67. **Principal Investigator**, *Rapidly Building High-Performance Information Agents*, AFOSR, \$780,000, December 2000.
68. **Principal Investigator**, *Active Template Reasoning and Optimization*, DARPA Active Templates Program, \$3,911,500, April 2000.
69. **Co-Principal Investigator**, *Agents for Plan Management*, DARPA Cooperative Agent-Based Systems Program, \$2,102,000, July 1998.
70. **Principal Investigator**, *Intelligent Agents for Retrieving, Filtering, and Managing Information*, AFOSR, \$985,000, October 1997.
71. **Co-Principal Investigator**, *A Portable SIMS System for Integrating Inconsistent and Partial Information Sources*, DARPA Intelligent Information Integration Program, \$1,867,000, October 1997.
72. **Co-Principal Investigator**, *Information Integration for Battlefield Awareness*, DARPA Battlefield Awareness and Data Dissemination Program, 750,000, June 1997.
73. **Co-Principal Investigator**, *SIMS for Dealmaker*, DARPA Advanced Logistics Program, \$544,000, August 1996.
74. **Co-Principal Investigator**, *SIMS II: Single Interface to Multiple Sources of Incomplete Information*, ARPA Intelligent Integration of Information Program, \$1,200,000, March 1994.

75. **Principal Investigator**, *Learning and Reformulation for Efficient Multidatabase Retrieval*, National Science Foundation Database and Expert Systems Program, \$259,000, September 1993.
76. **Principal Investigator**, *Building Information Servers*, DARPA Augmentation Award for Science and Engineering Research Training, \$139,000, March 1993.

Patents

• Geospatial Patents

- **Processing time-based geospatial data**
US 8,953,887
Inventors: Ching-Chien Chen, Craig A. Knoblock, Cyrus Shahabi
- **Precisely locating features on geospatial imagery**
US 8,675,995
Inventors: Ching-Chien Chen, Dipsy Kapoor, Craig A. Knoblock, Cyrus Shahabi
<https://patents.justia.com/patent/8675995>
- **Systems and methods for linking content to individual image features**
US 8,670,617
Inventors: Craig A. Knoblock, Cyrus Shahabi, Ching-Chien Chen, Dipsy Kapoor
<https://patents.justia.com/patent/8670617>
- **Dynamically linking relevant documents to regions of interest**
US 8,635,228
Inventors: Cyrus Shahabi, Craig A. Knoblock, Dipsy Kapoor, Ching-Chien Chen
<https://patents.justia.com/patent/8635228>
- **System and method for fusing vector data with imagery**
US 8,340,360
Inventors: Ching-Chien Chen, Dipsy Kapoor, Craig A. Knoblock, Cyrus Shahabi
<https://patents.justia.com/patent/8340360>
- **System and method for fusing geospatial data**
US 7,660,441
Inventors: Ching-Chien Chen, Craig A. Knoblock, Cyrus Shahabi, Yao-Yi Chiang
<https://patents.justia.com/patent/7660441>

• Predictive Pricing Patents

- **Performing predictive pricing based on historical data**
US 8,566,143
Inventors: Oren Etzioni, Alexander Yates, Craig A. Knoblock, Rattapoom Tuchinda
<https://patents.justia.com/patent/8566143>
- **Performing predictive pricing based on historical data**
US 7,974,863
Inventors: Oren Etzioni, Alexander Yates, Craig A. Knoblock, Rattapoom Tuchinda
<https://patents.justia.com/patent/7974863>
- **Performing predictive pricing based on historical data**
US 7,346,520
Inventors: Oren Etzioni, Alexander Yates, Craig A. Knoblock, Rattapoom Tuchinda
<https://patents.justia.com/patent/7346520>
- **Performing predictive pricing based on historical data**
US 7,010,494
Inventors: Oren Etzioni, Alexander Yates, Craig A. Knoblock, Rattapoom Tuchinda
<https://patents.justia.com/patent/7010494>

- **Wrapper Induction Patent**
 - **Wrapper induction by hierarchical data analysis**
US 6,606,625
Inventors: Ion Muslea, Steven Minton, Craig A. Knoblock
<https://patents.justia.com/patent/6606625>