MIT Sloan School of Management

MIT Sloan No. 4443-03
CISL No. 2003-08
February 2003

LIGHTS: Laboratory for Information Globalization and Harmonization Technologies and Studies

Nazli Choucri, Stuart Madnick, Michael Siegel, Richard Wang

LIGHTS:
Laboratory for Information Globalization and Harmonization Technologies and Studies

February 2003

Working Paper CISL# 2003-08

Nazli Choucri {nchoucri@mit.edu}
Stuart Madnick {smadnick@mit.edu}
Michael Siegel {msiegel@mit.edu}
Richard Wang {rwang@mit.edu}

Massachusetts Institute of Technology
Cambridge, Massachusetts

Laboratory for Information Globalization and Harmonization Technologies and Studies

Project Summary

Three important trends – unrelenting globalization, growing worldwide electronic connectivity, and increasing knowledge intensity of economic activity – are creating new opportunities for global politics, with challenging demands for information access, interpretation, provision and overall use. This has serious implications for two diverse domains of scholarship**: Information Technology** (IT) and **International Relations** (IR) in political science. Unless IT advances remain 'one step ahead' of such realities and complexities, strategies for better understanding and responding to emergent global challenges will be severely impeded. For example, the new Department of Homeland Security will rely on intelligence information from all over the world to develop strategic responses to a wide range of security threats. However, relevant information is stored throughout the world and by diverse agencies and in different media, formats, quality, and contexts. Intelligent integration of that information and improved modes of access and use are critical to developing policies designed to identify and anticipate sources of threat, to strengthen protection against threats on the United States, and to enhance the security of the nation.

The focus of this project is the creation of a **Laboratory for Information Globalization and Harmonization Technologies and Studies** (LIGHTS) which has two interrelated goals:

(1) Technologies: To research, design, develop, test, and implement the **System for Harmonized Information Processing** (SHIP*)* to facilitate access to, and correct interpretation of, essential globally distributed information that is critical to research and policy in the IR realm, as well as to other similarly complex domains

(2) Studies: To apply SHIP to important problems in International Relations in the complex domain of conflict and violence (dealing with emergent risks, threats, and uncertainties of global scale and scope), and special reference to (a) crises situations, (b) conflict and war; and (c) anticipation, monitoring and early warning.

The development of the SHIP system and related experiments in IR builds upon our prior work. The IT work builds on our research on the **COntext INterchange** (COIN) project, which focused on the integration of diverse distributed heterogeneous information sources (e.g., financial, supply chain, disaster relief) using ontologies, databases, context mediation algorithms, and wrapper technologies to overcome representational information conflicts. The political science IR work builds on our research on international conflict and on our experience with the **Global System for Sustainable Development** *(*GSSD), an Internet-based platform for information generation, provision, and integration of multiple domains, regions, languages, and epistemologies relevant to IR researchers.

Intellectual Merit: Although LIGHTS will leverage the results of these prior research projects, this will be the first research effort to effectively address ontological and temporal information conflicts as well as dramatically enhance information quality. These advances are needed to meet the ambitious and important goals of the LIGHTS-IR studies. Addressing problems of national and global interests in such rapidly changing complex environments require the use of observations from disparate sources, using different interpretations, at different times, for different purposes, with different biases, and for a wide range of different uses and users. The SHIP research will focus on integrating and enhancing information both over individual domains (i.e., a single shared ontology) and across multiple domains. A core innovation is reflected in the notion of a Collaborative Domain Space (CDS), within which applications in a common domain can share, analyze, modify, and develop information. Applications also can span multiple domains via Linked CDSs. The SHIP will provide a novel basis for actionable domain knowledge representation and include the reasoning algorithms required for processing over a range of heterogeneous sources and applications. The PIs have considerable experience with these research areas and the organization and management of such large scale, international, and diverse research projects.

Multi-Disciplinary and Diversity: The PIs come from three different Schools at MIT: Management, Engineering, and Humanities, Arts & Social Sciences. The faculty and graduate students come from about a dozen nationalities and diverse ethnic, racial, and religious backgrounds. The currently identified external collaborators come from over 20 different organizations and many different countries, industrial as well as developing.

Broader impacts from the Research: The anticipated results will apply to any complex domain that relies on heterogeneous distributed data to address and resolve compelling problems. This initiative is supported by international collaborators from (a) scientific and research institutions, (b) business and industry, and (c) national and international agencies. Expected research products include: SHIP, a software platform; an IR-based knowledge repository; and diverse applications in research and education which are anticipated to significantly impact the way complex organizations, and society in general, understand and manage critical global challenges. The research results will be widely disseminated both through scholarly publications as well as new teaching materials, including delivery through innovative channels, such as MIT's OpenCourseware initiative.

Section 1. Project Overview and Significance

1.1 Emergent Challenges to Global Information

The convergence of three distinct but interconnected trends – unrelenting globalization, growing world-wide electronic connectivity, and increasing knowledge intensity of economic activity – is creating powerful new parameters for global politics which are reflected in critical new challenges to current modes of information access and understanding. First, the discovery and retrieval of relevant information has become a daunting task due to the sheer volume, scale, and scope of information on the Internet, its geographical dispersion, varying context, heterogeneous sources, and variable quality. Second, the opportunities presented by this transformation are shaping new demands for improved information generation, management, and analysis. Third, more specifically, the increasing diversity of Internet uses and users points to the importance of cultural and contextual dimensions of information and communication. There are significant opportunity costs associated with overlooking these challenges, potentially hindering both empirical analysis and theoretical inquiry so central to many scholarly disciplines, and their contributions to national policy. *This proposal seeks to identify new ways of addressing these challenges by significantly improving access to diverse, distributed, and disconnected sources of information.* Although this effort will focus on the realm of International Relations, the results will have relevancy to the broader field of Political Science (and related social sciences), as well as to most scientific endeavors that have such information needs.

1.2 Relevance to Political Science Scholarship

Political science is generally understood to be the systematic study of '*who* gets *what*, *when*, and *how*' [Las58], which translates roughly into actor (or agent), stake (goal or utility), timing, and strategy (action, behavior). In this field, the demands for information in a rapidly changing world surpass existing capabilities for information access, retrieval, organization, interpretation, and use – thus creating (a) *gaps* between needs and capabilities, (b) *lags* between the availability of information and its access for effective use in scholarly as well as policy-relevant research, and (c) *barriers* to effective use created by disconnects across format, sources, language, cultural differences, and contextual conditions. International relations (IR), a sub-field of political science, focuses on the international domain (and interfaces with domestic politics), and examines issues such as sovereignty, security, cross-border conflicts, and modes of cooperation. The gaps, lags, and barriers described above are especially pertinent to theoretically- driven and empirically- informed inquiries in the IR field.

1.2.1 International Relations (IR) Examples

This project will focus on information needs in the conflict realm of international relations, involving emergent risks, threats of varying intensity, and uncertainties of potentially global scale and scope. Specifically, we propose to focus on: (a) crisis situations; (b) conflicts and war; and (c) anticipation, monitoring, and early warning. Information needs for research in these domains are extensive and vary depending on: (1) the *salience* of information (i.e. the criticality of the issue), (2) the *extent of customization*, and (3) the *complexity* at hand. More specifically, in:

- **Crisis situations:** the needs are characteristically immediate, usually highly customized, and generally require complex analysis, integration, and manipulation of information. International crises are now impinging more directly than ever before on national security, thus rendering the information needs and requirements even more pressing.
- **Conflicts and War**: the needs are not necessarily time-critical, are customized to a certain relevant extent, and involve a multifaceted examination of information. Increasingly, it appears that coordination of information access and analysis across a diverse set of players (or institutions) with differing needs and requirements (perhaps even mandates) is more the rule rather than the exception in cases of conflict and war.
- **Anticipation, Monitoring and Early Warning**: the needs tend to be gradual, involve routinized searches, but require extraction of information from sources that may evolve and change over time. Furthermore, in today's global context, 'preventative action' may even take on new urgency, and create new demands for information services.

The examples in Table 1 illustrate the types of information needs required for effective research, education, decision-making, and policy analysis on a range of conflict issues for which there is considerable scholarship in place. These issues remain central to matters of security in this increasingly globalized world.

Illustrative Cases	Example of Information Needs	Intended Use of Information
1. Strategic Requirements for Managing Cross-Border Pressures in a *Crisis* The UNHCR needs to respond to the dislocation and large numbers of Afghans into neighboring countries, triggered by war in Afghanistan.	Logistical and infrastructure information for setting up refugee camps, such as potential sites, sanitation, and potable water supplies.	Facilitated coordination of relief agencies with up-to-date information during a crisis for more rapid response (as close to real time as possible).
2. Capabilities for Management during an Ongoing *Conflict & War* The goal of the newly established UNEP-Balkans group is to assess whether the ongoing Balkan conflict has had significant environmental and economic impacts on the region. The data, extensive as it may be, is dispersed and presented in different contexts.	Environmental and economic data on the region prior to the initiation/ escalation of the conflict. Comparison of this data with newly collected data to assess the impacts to environmental and economic viability.	Improved decision making during conflicts and war - taking into account contending views and changing strategic conditions - in order to better prepare for, and manage, future developments and modes of resolution.
3. Strategic Response to Security Threats for *Anticipation, Prevention, and Early Warning* The newly-created Department of Homeland Security needs to coordinate U.S. government efforts with foreign governments using information from different regions of the world.	Intelligence data from foreign governments, non-governmental agencies, US agencies, and leading opinion leaders worldwide.	Streamline potentially conflicting information content and sources in order to facilitate coherent anticipation, preventive monitoring, and early warning.

Table 1. Illustrating Information Needs in Three Contexts

1.3 Addressing Information Needs in the Conflict Realm

1.3.1 Operational Example

For illustrative purposes only, this section elaborates on the *gaps, lags*, and *barriers*, described above, which are prevalent in the types of examples illustrated by Example 2 in Table 1. The specific question is: **to what extent have economic performance and environmental conditions in Yugoslavia been affected by the conflicts in the region?** The answer to this question could shape policy priorities for different national and international institutions, as well as reconstruction strategies, and may even determine which agencies will be the leading players. Moreover, there are potentials for resumed violence and the region's relevance to overall European stability remains central to the US national interest. This is not an isolated case, by any means, but one that illustrates concurrent challenges for information compilation, analysis, and interpretation – under changing conditions.

For example, if we are interested in determining the change of carbon dioxide (CO_2) emissions in the region, normalized against the change in GDP - before and after the outbreak of the hostilities – we need to take into account territorial and jurisdictional boundaries, changes in accounting and recording norms, and varying degrees of autonomy. User requirements add another layer of complexity. For example, what units of CO_2 emissions and GDP should be displayed, and what unit conversions need to be made from the information sources? Which Yugoslavia is of concern to the user: the country defined by its current borders, or the entire geographic area formerly known as Yugoslavia? One of the effects of the war is that the region, which used to be one country consisting of six republics and two provinces, has subsequently been reconstituted into five legal entities (countries), each having its own reporting formats, currency, units of measure, and new socio-economic parameters. In other words, the meaning of the request for information will differ, depending on the *actors*, *actions, stakes* and *strategies* involved.

In this simple case, we suppose that the request comes from a reconstruction agency interested in the following values: CO_2 emission amounts (in tons/yr), CO_2 per capita, annual GDP (in million USD/yr), GDP per capita, and the ratio CO_2/GDP (in tons CO_2/million USD) for the entire region of the former Yugoslavia (see the alternative User 2 scenario in Table 2). A restatement of the question would then become: **what is the change in CO_2 emissions and GDP in the region formerly known as Yugoslavia before and after the war?**

1.3.2 Diverse Sources and Contexts

By necessity, to answer this question, one needs to draw data from diverse types of sources (we call these differing *domains* of information) - such as, economic data (e.g., the World Bank, UN Statistics Division), environmental data (e.g., Oak Ridge National Laboratory, World Resources Institute), and country history data (e.g., the CIA Factbook), as illustrated in Table 2 below. Merely combining the numbers from the various sources is likely to produce serious errors due to different sets of assumptions driving the representation of the information in the sources These assumptions are often not explicit but are an important representation of 'reality' (we call these the

meaning or *context* of the information, which will be explained in more detail in Section 2.)

<table>
<tr><th>Domain and Sources Consulted</th><th>Sample Data Available</th><th>Basic Question, Information User Type & Usage</th></tr>
<tr><td>Economic Performance
• World Bank's World Development Indicators database
• UN Statistics Division's database
• Statistics Bureaus of individual counties</td><td>A. Annual GDP and Population Data:
<table><tr><th>Country</th><th>T0.GDP</th><th>T0.Pop</th><th>T1.GDP</th><th>T1.Pop</th></tr><tr><td>YUG</td><td>698.3</td><td>23.7</td><td>1627.8</td><td>10.6</td></tr><tr><td>BIH</td><td></td><td></td><td>13.6</td><td>3.9</td></tr><tr><td>HRV</td><td></td><td></td><td>266.9</td><td>4.5</td></tr><tr><td>MKD</td><td></td><td></td><td>608.7</td><td>2.0</td></tr><tr><td>SVN</td><td></td><td></td><td>7162</td><td>2.0</td></tr></table>- GDP in billions local currency per year
- Population in millions</td><td rowspan="4">Question:
How did economic output and environmental conditions change in YUG over time?

User 1: YUG as a geographic region bounded at T0:
<table><tr><td>Parameter</td><td>T0</td><td>T1</td></tr><tr><td>CO_2</td><td>35604</td><td>29523</td></tr><tr><td>CO_2/capita</td><td>1.50</td><td>1.28</td></tr><tr><td>GDP</td><td>66.5</td><td>104.8</td></tr><tr><td>GDP/capita</td><td>2.8</td><td>4.56</td></tr><tr><td>CO_2/GDP</td><td>535</td><td>282</td></tr></table>User 2: YUG as a legal, autonomous state
<table><tr><td>Parameter</td><td>T0</td><td>T1</td></tr><tr><td>CO_2</td><td>35604</td><td>15480</td></tr><tr><td>CO_2/capita</td><td>1.50</td><td>1.46</td></tr><tr><td>GDP</td><td>66.5</td><td>24.2</td></tr><tr><td>GDP/capita</td><td>2.8</td><td>1.1</td></tr><tr><td>CO_2/GDP</td><td>535</td><td>640</td></tr></table>Note:

T0: 1990 (prior to breakup)
T1: 2000 (after breakup)
CO_2: 1000's tons per year
CO_2/capita: tons per person
GDP: billions USD per year
GDP/capita: 1000's USD per person
CO_2/GDP: tons per million USD</td></tr>
<tr><td>Environmental Impacts
• Oak Ridge National Laboratory's CDIAC database
• WRI database
• GSSD
• EPA of individual countries</td><td>B. Emissions Data:
<table><tr><th>Country</th><th>T0</th><th>T1</th></tr><tr><td>YUG</td><td>35604</td><td>15480</td></tr><tr><td>BIH</td><td></td><td>1279</td></tr><tr><td>HRV</td><td></td><td>5405</td></tr><tr><td>MKD</td><td></td><td>3378</td></tr><tr><td>SVN</td><td></td><td>3981</td></tr></table>- Emissions in 1000s tons per year</td></tr>
<tr><td>Country History:
• CIA
• GSSD</td><td>T0.{YUG} = T1.{YUG, BIH, HRV, MKD, SVN}
(i.e., geographically, YUG at T0 is equivalent to YUG+BIH+HRV+MKD+SVN at T1)</td></tr>
<tr><td>Mappings Defined:
• Country code
• Currency code
• Historical exchange rates*

* Note: Hyperinflation in YUG resulted in establishment of a new currency unit in June 1993. Therefore, T1.YUN is completely different from</td><td><table><tr><th>Country</th><th>Code</th><th>Currency</th><th>Currency Code</th></tr><tr><td>Yugoslavia</td><td>YUG</td><td>New Yugoslavian Dinar</td><td>YUN</td></tr><tr><td>Bosnia and Herzegovia</td><td>BIH</td><td>Marka</td><td>BAM</td></tr><tr><td>Croatia</td><td>HRV</td><td>Kuna</td><td>HRK</td></tr><tr><td>Macedonia</td><td>MKD</td><td>Denar</td><td>MKD</td></tr><tr><td>Slovenia</td><td>SVN</td><td>Tolar</td><td>SIT</td></tr></table><table><tr><th>C_From</th><th>C_To</th><th>T0</th><th>T1</th></tr><tr><td>USD</td><td>YUN</td><td>10.5</td><td>67.267</td></tr><tr><td>USD</td><td>BAM</td><td></td><td>2.086</td></tr><tr><td>USD</td><td>HRK</td><td></td><td>8.089</td></tr><tr><td>USD</td><td>MKD</td><td></td><td>64.757</td></tr><tr><td>USD</td><td>SIT</td><td></td><td>225.93</td></tr></table></td></tr>
</table>

Table 2. Operational Example: Information Needs in Cases of Conflict

The purpose of Table 2 is to illustrate some of the complexities in a seemingly simple question. In addition to variations in data sources and domains, there are significant differences in contexts and formats, critical temporality issues, and data conversions that all factor into the user's information needs. As specified in the table, time T0 refers to a date b*efore the war* (e.g., 1990), when the entire region was a single country (referred to as "YUG"). Time T1 refers to a date a*fter the war* (e.g., 2000), when the country "YUG" retains its name, but has lost four of its provinces, which are now independent countries. The first column of Table 2 lists some of the sources and domains covered by this question. The second column shows sample data that could be extracted from the sources. The bottom row of this table lists auxiliary mapping information that is needed to understand the meanings of symbols used in the other data sources. For example, when the GDP for Yugoslavia is written in YUN units, a currency code source is needed to understand that this symbol represents the Yugoslavian Dinar. The third column

lists the outputs and units requested by the user. Accordingly, for User 1, a simple calculation based on data from country "YUG" will invariably give a wrong answer. For example, deriving the CO_2/GDP ratio by simply summing up the CO_2 emissions and dividing it by the sum of GDP from sources A and B will not provide a correct answer.

1.3.3 Manual Approach

Given the types of data shown in Table 2, along with the appropriate context knowledge (some of which is shown in italics), an analyst could determine the answer to our question. The proper calculation involves numerous steps, including selecting the necessary sources, making the appropriate conversions, and using the correct calculations. For example:

For time T0:

1. Get CO_2 emissions data for "YUG" from source B;
2. Convert it to tons/year using scale factor 1000; call the result X;
3. Get GDP data from source A;
4. Convert to USD by looking up currency conversion table, an auxiliary source; call the result Y;
5. No need to convert the scale for GDP because the receiver uses the same scale, namely, 1,000,000;
6. Compute X/Y (equal to 535 tons/million USD in Table 2).

For time T1:

1. Consult source for country history and find all countries in the area of former YUG;
2. Get CO_2 emissions data for "YUG" from source B (or a new source);
3. Convert it to tons/year using scale factor 1000; call the result X1;
4. Get CO_2 emissions data for "BIH" from source B (or a new source);
5. Convert it to tons/year using scale factor 1000; call the result X2;
6. Continue this process for the rest of the sources to get the emissions data for the rest of the countries;
7. Sum X1, X2, X3, etc. and call it X;
8. Get GDP for "YUG" from source A (or alternative); Convert it to USD using the auxiliary sources;
9. No need to convert the scale factor; call the result Y1;
10. Get GDP for "BIH" from source E; Convert it to USD using the auxiliary sources; call the result Y2;
11. Continue this process for the rest of the sources to get the GDP data for the rest of the countries;
12. Sum Y1, Y2, Y3, etc. and call it Y;
13. Compute X/Y (equal to 282 tons/million USD in Table 2).

The complexity of this task would be easily magnified if, for example, the CO_2 emissions data from the various sources were all in different metrics or, alternatively, if demographic variables were drawn from different institutional contexts (e.g., with or without counting refugees). This example shows some of the operational challenges if a user were to manually attempt to answer this question. This case highlights just some of the common data difficulties in the IR domain where information reconciliation continues to be made 'by hand'. It is easy to see why such analysis can be very labor intensive and error-prone. This makes it difficult under "normal" circumstances and possibly impossible under time-critical circumstances.

1.3.4 SHIP: A Better Way

A key goal of this research effort is to create the System for Harmonized Information Processing (SHIP), a system that can automatically determine and reliably perform the steps shown above in response to a user's request. SHIP will be capable of storing the necessary context information about the sources and users and have a reasoning engine capable of determining the necessary sources, conversions, and calculations necessary. The COIN system, to be described more fully later, has proven the feasibility of this approach in more limited situations. SHIP will be the next generation

1.4 Impact on Current IR Practices and Other Scientific Domains

Currently, IR theory isolates behaviors of agents (states, firms, etc.) from their environmental context (i.e., by separating social and natural systems). As a result, there is little theoretical guidance for modes of interface between activities and environments, and only nominal practice in integrating information on incidences of conflict with environmental data. This is a case where actual practice has preceded theory and analysis. For example, leading IR scholars on causes and consequences of war (notably [Van99]) ignore 'environment' factors entirely, and only recently have leading historians addressed potential interconnections [McN00]. Barriers to information access and interface, as shown above, have hampered efforts to draw effectively upon indices of natural systems and social systems, and impede serious efforts to develop more integrated perspectives on the matter of actors, actions, and outcomes – particularly in cases where conflict and violence create interdependent social and environmental

dislocations. SHIP can significantly improve current IR practices by making it feasible and practical to make better use of the increasingly available, though highly diverse, sources and types of information. Our purpose is to produce useful tools and a broad architecture that can be applied to many scholarly disciplines, by focusing on information needs that are strategically compelling, emerge from complex domains, require cross-disciplinary connectivity and linkages, and involve a diversity of actors and agents, distributed data sources, and forms of provision.

1.5 Existing Foundations – COIN and GSSD

Important research in two areas has already been completed that provides important foundations for addressing the emergent challenges discussed in Table 1: the *COntext INterchange* Project (COIN, see context2.mit.edu) and the *Global System for Sustainable Development* (GSSD, see gssd.mit.edu).

1.5.1 COIN

The *COntext INterchange* (COIN) Project has developed a basic theory, architecture, and software prototype for supporting intelligent information integration employing context mediation technology [MAD99, GBM*99, GoBM96, Goh96, SM91a]. We propose to utilize the foundation of COIN to develop theories and methodologies for the System for Harmonized Information Processing. The fundamental concept underlying such a system is the representation of knowledge as **Collaborative Domain Spaces (CDSs)**. A **CDS** is a grouping of the knowledge including source schemas, data context, conversion functions, and source capabilities as related to a single domain ontology. The software components needed to provide harmonized information processing (i.e. through the use of a CDS or collections of linked CDSs) include a context mediation engine [BGL*00, Goh96], one or more ontology library systems, a context domain and conversion function management system, and a query execution and planner [Fynn97]. In addition, support tools are required to allow for applications' (i.e. receivers') context definition and source definitions to be added and removed easily (i.e., schemas, contexts, capabilities). Developing a flexible, scalable software platform will require significant additional research in a number of key research areas as described in Section 2.4.

1.5.2 GSSD

The *Global System for Sustainable Development* serves as an Internet-based platform for exploring forms of information access, provision, and integration across multiple information sources, languages, cultural contexts, and ontologies. GSSD has an extensive, quality-controlled set of ontologies related to system sustainability, which is relevant to the field of international relations. In addition, GSSD has made considerable gains into understanding the organization and management of large scale, distributed, and diverse research teams, including cross-national (China and Japan, and countries in the Middle East and Europe) and institutional partners (private, public, and international agencies). Designed and implemented by social scientists, GSSD is seen as demonstrating 'opportunities for collaboration and new technologies,' according to the National Academy of Engineering [RAC01, p. viii]. GSSD databases cover issues related to dynamics of conflict, as well as other domains relevant to our proposed research, such as migration, refugees, unmet human needs, as well as evolving efforts at coordinated international actions. GSSD provides a rich testing ground for the new information technologies we propose to develop such as automated methods for information aggregation from various sources, context mediation capabilities, customized information retrieval capabilities, and ontology representations.

1.6. Research Team

Due to the multi-disciplinary nature of this project, we have composed a research team that is uniquely qualified to conduct this work. The PIs of this project come from MIT's School of Humanities, Arts, and Social Sciences (Choucri), School of Engineering (Madnick and Wang), and School of Management (Siegel and Madnick), and the students who will contribute significantly to the research come from all these diverse Schools. Furthermore, the PIs have extensive research experience in the critical areas necessary to accomplish the goals of this project.

1.7. Proposal Organization

The remainder of this proposal will elaborate on the intended research tasks. Section 2 will describe research needs in Information Technology and how these capabilities can benefit International Relations in Section 3. Section 4 provides a brief description of the new laboratory that will ensure coherence among the components of the project and also handle outreach activities. Finally Sections 5 and 6 will present the anticipated contributions of the project, with a focus on educational impacts.

Section 2. Information Technology Research

2.1 Needs for Harmonized Information Processing and Collaborative Domain Spaces

Advances in computing and networking technologies now allow extensive volumes of data to be gathered, organized, and shared on an unprecedented scale and scope. Unfortunately, these newfound capabilities by themselves are only marginally useful if the information cannot be easily **extracted** and **gathered** from **disparate sources**, if the information is represented with **different interpretations**, and if it must satisfy **differing user needs** [MHR00, MAD99, CFM*01]. The data requirements (e.g., scope, timing) and the sources of the data (e.g., government, industry, global organizations) are extremely diverse. The need for intelligent harmonization of heterogeneous information is important to all information-intensive endeavors – which encompasses many disciplines and issue-areas, including governments, education, science and engineering. The fundamental technology research to be performed has broad relevancy for all global applications, such as Manufacturing (e.g., Integrated Supply Chain Management), Transportation/Logistics (e.g., In-Transit Visibility), Government/Military (e.g., Total Asset Visibility), and Financial Services (e.g., Global Risk Management). It is proposed that the application focus for this research effort be in the domain of international relations that, by definition, takes into account internal as well as external dimensions of relations among actors in both the public and the private domains.

This research effort will:

1. Analyze the data and technology requirements for the categories of problems described in Section 1;
2. Research, design, develop and test extensions and improvements to the underlying COIN and GSSD theory and components;
3. Provide a scalable, flexible platform for servicing the range of applications described in Section 1; and
4. Demonstrate the effectiveness of the theories, tools, and methodologies through technology transfer to other collaborating organizations.

2.2 Illustrative Example of Information Extraction, Dissemination, and Interpretation Challenges

As an illustration of the problems created by information disparities, let us refer back to the example from the conflict realm introduced in Section 1.3. The question was**: what are the impacts of CO_2 emissions on economic performance in Yugoslavia.** It is necessary to draw data from diverse sources such as CIA Worldbook (for current boundaries), World Resources Institute (for CO_2 emissions), and the World Bank (for economic data).

There are many additional information challenges that had not been explicitly noted earlier, such as:

Information Extraction: Some of the sources may be full relational databases, in which case there is the issue of remote access. In many other cases, the sources may be traditional HTML web sites, which are fine for viewing from a browser but not effective for combining data or performing calculations (other than manually "cut & paste"). Other sources might be tables in a text file, Word document, or even a spreadsheet. Although the increasing use of eXtensible Markup Language (XML) will reduce some of these interchange problems [MAD01], we will continue to live in a very heterogeneous world for quite a while to come. So we must be able to extract information from all types of sources.

Information Dissemination: The users want to use the resulting "answers" in many ways. Some will want to see the desired information displayed in their web browser but others might want the answers to be deposited into a database, spreadsheet, or application program for further processing.

Information Interpretation: Although the problems of information extraction and dissemination will be addressed in this research, the most difficult challenges involve information interpretation, as illustrated below.

Specifically, an example question is: "What is the change of CO_2 emissions per GDP in Yugoslavia before and after the Balkans war?"

Before the war (time T0), the entire region was one country. Data for CO_2 emissions was in thousands of tons/year, and GDP was in billions of Yugoslavian Dinars. *After the war* (time T1), Yugoslavia only has two of its original five provinces; the other three provinces are now four independent countries, each with its own currency. The size and population of the country, now known as Yugoslavia, has changed. Even Yugoslavia has introduced a new currency to combat hyperinflation.

From the perspective of any one agency, **UNEP** for example, the question: "How have CO_2 emission per GDP changed in Yugoslavia after the war?" may have multiple interpretations. Not only does each source have a context, but so does each user (also referred to as a receiver). For example, does the user mean Yugoslavia as the original geographic area (depicted as *user 1* in Table 2) or as the legal entity, which has changed size (*user 2*). To answer the question correctly, we have to use the changing context information. A simple calculation based on the "raw" data will not give the right answer. As seen earlier, the calculation will involve many steps, including selecting necessary sources, making appropriate conversions, and using correct calculations. Furthermore, each user

might have a different preferred context for their answer, such as: tons/million USD or kilograms/billion EURO, etc. More of these information harmonization challenges will be highlighted in Section 2.4.

Although seemingly simple, this example addresses some of the most complex issues in IR: namely the impact of changing legal jurisdictions and sovereignties on (a) state performance, (b) salience of socio-political stress, (c) demographic shifts and (d) estimates of economic activity, as critical variables of note. Extending this example to the case of the former Soviet Republics, before and after independence, is conceptually the same type of challenge – with greater complexity. For example, the US Department of Defense is interested in demographic distributions around oil fields (by ethnic group) and before and after independence. Alternatively, UNEP is interested in CO_2 emissions per capita given that these are oil-producing regions. On the other hand, foreign investors will be interested in insurance rates before and after independence.

The information shown as footnotes in Table 2 (e.g., "Population in millions") illustrates **context knowledge**. Sometimes this context knowledge is explicitly provided with the source data (but still must be accessed and processed), but many times it must be found in other sources, and on occasion someone must be asked to track down and explain the meaning of the data. The good news is that such context knowledge almost always exists, but it is often widely **distributed** within and across organizations. Thus, a central focus of this part of the effort is to support **the acquisition, organization, and effective intelligent usage of distributed context knowledge to support information harmonization and collaborative domains**.

2.3 Research Platform

The MIT COntext INterchange (COIN) Project has developed a platform including a theory, architecture, and basic prototype for such intelligent harmonized information processing. COIN is based on database theory and mediators [Wied92, Wied99]. Context Interchange is a mediation approach for semantic integration of disparate (heterogeneous and distributed) information sources as described in [BGL*00 and GBM*99]. The Context Interchange approach includes not only the mediation infrastructure and services, but also wrapping technology and middleware services for accessing the source information and facilitating the integration of the mediated results into end-users applications (see Figure 1).

The wrappers are physical and logical gateways providing uniform access to the disparate sources over the network [Chen99, FMS00a, FMS00b]. The set of Context Mediation Services, comprises a Context Mediator, a Query Optimizer and a Query Executioner. The Context Mediator is in charge of the identification and resolution of potential semantic conflicts induced by a query. This automatic detection and reconciliation of conflicts present in different information sources is made possible by ontological knowledge of the underlying application domain, as well as informational content and implicit assumptions associated with the receivers and sources.

The result of the mediation is a mediated query. To retrieve the data from the disparate information sources, the mediated query is then transformed into a query execution plan, which is optimized, taking into account the topology of the network of sources and their capabilities. The plan is then executed to retrieve the data from the various sources, then results are composed and sent to the receiver.

In a heterogeneous and distributed environment, the mediator transforms a query written in terms known in the user or application program context (i.e., according to the user's or programmer's assumptions and knowledge) into one or more queries in the terms of the component sources. The individual subqueries may still involve several sources. However, subsequent planning, optimization and execution phases are needed [AKS96, Fynn97]. The planning and execution phases must consider the limitations of the sources and the topology and costs of the network (especially when dealing with non-database sources, such as web pages or web services). The execution phase is in charge of the scheduling of the query execution plan and the realization of the complementary operations that could not be handled by the sources individually (e.g. a join across sources).

Where a large number of independent information sources are accessed (as is now possible with the global Internet), flexibility, scalability, and non-intrusiveness will be of primary importance. Traditional tight-coupling approaches to semantic interoperability rely on the *a priori* creation of federated views on the heterogeneous information sources. These approaches do not scale-up efficiently given the complexity involved in constructing and maintaining a shared schema for a large number of, possibly independently managed and evolving, sources. Loose-coupling approaches rely on the user's intimate knowledge of the semantic conflicts between the sources and the conflict resolution procedures. This reliance becomes a drawback for scalability when this knowledge grows and changes as more sources join the system and when sources are changing. The Context Interchange (COIN) approach is a middle ground between these two approaches. It allows queries to the sources to be mediated, i.e. semantic conflicts to be identified and solved by a context mediator through comparison of contexts associated with the sources and receivers concerned by the queries. It only requires the minimum adoption of a common Domain Model that defines the domain of discourse of the application.

The knowledge needed for harmonization is formally modeled in a COIN framework [Goh96], The COIN framework is a mathematical structure offering a robust foundation for the realization of the Context Interchange strategy. The COIN framework comprises a data model and a language, called COINL, of the Frame-Logic (F-Logic) family [KLW95, DT95]. The framework is used to define the different elements needed to implement the strategy in a given application:

- The Domain Model is a collection of rich types (semantic types) defining the domain of discourse for the integration strategy;
- Elevation Axioms for each source identify the semantic objects (instances of semantic types) corresponding to source data elements and define integrity constraints specifying general properties of the sources;
- Context Definitions define the different interpretations of the semantic objects in the different sources or from a receiver's point of view.

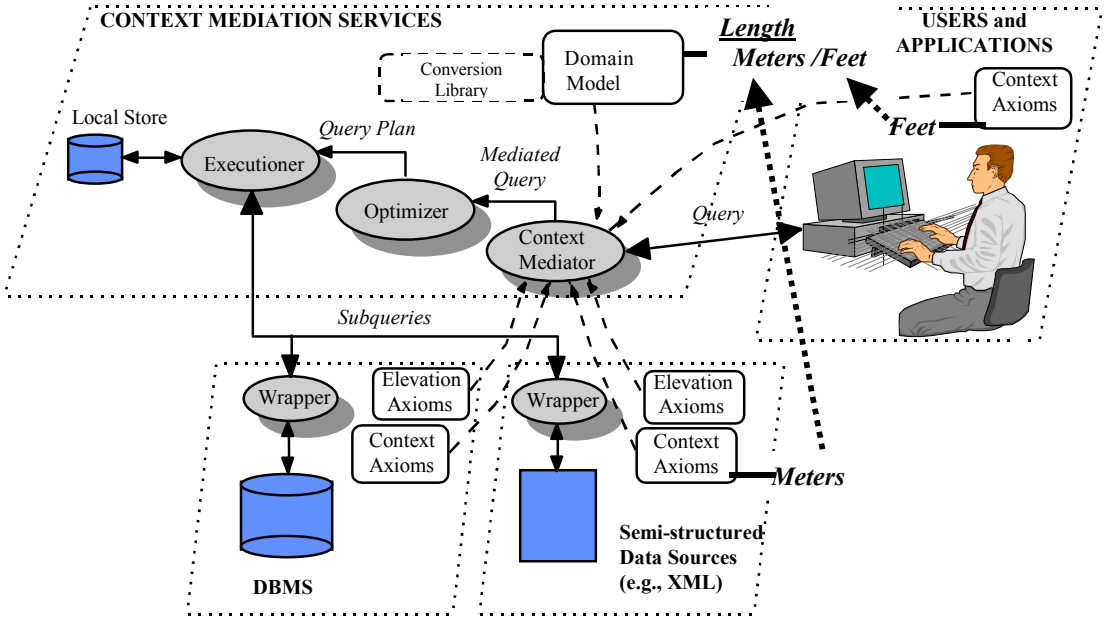

Figure 1. The Architecture of the Context Interchange System

The comparison and conversion procedure itself is inspired by the Abductive Logic Programming framework [KKT93] and can be qualified as an abduction procedure, to take advantage of its formal logical framework. One of the main advantages of the abductive logic programming framework is the simplicity in which it can be used to formally combine and to implement features of query processing, semantic query optimization and constraint programming.

2.4. Research Tasks and Expected Contributions

Although the existing COIN system and its underlying research provides a powerful "head start" and platform for harmonized information processing, it is still inadequate to address the types of needs illustrated in Section 1. Performing the important research to address these additional requirements and to produce our System for Harmonized Information Processing is the primary focus of this part of the effort.

1. Extended Domain of Knowledge – Equational and Temporal Context. In addition to the types of domain and context knowledge currently handled by the COIN framework, we need to perform research to add capabilities for both the representation and reasoning to provide support for equational [FGM02] and temporal context. Equational context refers to the knowledge such as "average GDP per person (AGDP)" means "total GDP" divided by "population." In some data sources, AGDP explicitly exists (possibly with differing names and in differing units), but in other cases it may not explicitly exist but could be calculated by using "total GDP" and "population" from one or more sources – if that knowledge existed and was used effectively. Temporal context refers to the fact that context not only varies across sources but also across time. Thus, the implied currency context for France's GDP prior to 2002 might be French Francs but after 2002 it is in Euros. If one were performing a longitudinal study over multiple years from multiple sources, it is important that this variation in context over time be understood and processed appropriately. In International Relations (IR), a seemingly straightforward variable

like the size of 'military expenditures' across countries is defined differently depending on the rules of inclusion or exclusion (as, for example, of military pensions) used in different jurisdictions.

2. Linked Collaborative Domain Spaces. The existing COIN framework provides representation and reasoning capabilities for a single domain. Although there are a number of ontology library systems that allow for management of multiple ontologies [DSW*99, DFen01 Fensel01, HelfH00], they have limitations in scalability and dynamically incorporating new ontological knowledge. Especially, they lack the capability of representing rich context knowledge needed for reconciling differences among sources. The primary focus of this overall research effort, and driven by the international relations setting, is the ability to operate in a multi-disciplinary environment across multiple linked collaborative domain spaces. The representational capabilities to relate concepts across domains, and efficiently maintain the effectiveness of these collaborative domain spaces is critically important – especially in an environment where we believe the underlying domains themselves will continually undergo evolution. In IR, for some users, the reality of domain shifts itself is the defining feature of interest [Nuna01].

3. Advanced Mediation Reasoning and Services. The COIN abductive framework can also be extrapolated to problem areas such as integrity management, view updates and intentional updates for databases [Chu00]. Because of the clear separation between the declarative definition of the logic of mediation into the COINL program from the generic abductive procedure for query mediation, we are able to adapt our mediation procedure to new situations such as mediated consistency management across disparate sources, mediated update management of one or more database using heterogeneous external auxiliary information or mediated monitoring of changes. The COIN approach holds the knowledge of the semantics of data in each context and across contexts in declarative logical statements separate from the mediation procedure. An update asserts that certain data objects must be made to have certain values in the updater's context. By combining the update assertions with the COIN logical formulation of context semantics, we can determine whether the update is unambiguous and feasible, and if so, what source data updates must be made to achieve the intended results. If ambiguous or otherwise infeasible, the logical representation may be able to indicate what additional constraints would clarify the updater's intention sufficiently for the update to proceed. We will build upon the formal system underlying our current framework, F-Logic and abductive reasoning, and extend the expressiveness and the reasoning capabilities leveraging ideas developed in different yet similar frameworks such as Description Logic and classification.

4. Automatic Source Selection. A natural extension is to leverage context knowledge to achieve context-based automatic source selection. One particular kind of context knowledge useful to enable automatic source selection is the content scope of data sources [TM98]. Data sources differ either significantly or subtly in their coverage scopes. In a highly diverse environment with hundreds and thousands of data sources, differences of content scopes can be valuably used to facilitate effective and efficient data source selection. Integrity constraints in COINL and the consistency checking component of the abductive procedure provide the basic ingredients to characterize the scope of information available from each source, to efficiently rule out irrelevant data sources and thereby speed up the selection process. For example, a query requesting information about *companies with assets lower than $2 million* can avoid accessing a particular source based on knowledge of integrity constraints stating that *the source only reports information about companies listed in the New York Stock Exchange* (NYSE), and that *companies must have assets larger than $10 million to be listed in the NYSE*. In general, integrity constraints express necessary conditions imposed on data. However, more generally, a notion of completeness degree of the domain of the source with respect to the constraint captures a richer semantic information and allows more powerful source selection. For instance, a source could contain exactly or at least all the data verifying the constraint (e.g., all the companies listed in the NYSE are reported in the source). In the IR domain, the source may be influenced by institutional objectives, resulting in major differences in metrics (for concepts like 'terrorism') due to differences in definitions of the concept itself.

5. Source Quality. Not only do the sources vary in semantic meaning, they also vary in quality. We must be able to represent and reason about the quality attributes of the sources. Although there has been some basic research on modeling the semantics of data quality [WKM93], significant additional research must be done to advance and formalize these notions and then incorporate them into the SHIP system

6. Attribution Knowledge Processing. For quality assessment and other reasons, it is important to know the attribution of the sources [LCN*99, LMB98]. For example, it can be important to know that although three different sources agree on a controversial piece of the information (e.g., casualties in the Afghanistan war), all three sources acquired that information from the same, maybe questionable, origin source. Thus, attribution metadata must be represented and processed in our system.

7. Domain Knowledge Processing – Improving Computer Performance. While domain and context knowledge processing has been shown to have considerable conceptual value [CZ98, MBM*98, LMS96b, SW92], its application in real situations requires both efficiency and scalability across large numbers of sources, quantities

and kinds of data, and demand for services. The scalability and optimization of this mediation processing for large numbers of sources across multiple collaborative domains and contexts will invariably be important.

8. Domain Knowledge Acquisition – Improving Human Performance. Domain and context knowledge acquisition are also very important. One essential property to be emphasized is the independence of the domains and sources. Our approach is non intrusive and respects their independence (i.e. autonomy). To effectively use the expressive power of the constructs and mechanisms in COINL, it is important that the human knowledge sources be able to easily provide the needed domain and context knowledge. It is therefore essential to develop an appropriate flexible methodology, and the tools supporting this methodology.

9. Operational System for Harmonized Information Processing A critical goal of this project is to develop a fully operational System for Harmonized Information Processing that will be used to support the types of IR inquiries and challenges listed in Section 1. It is essential that this system be developed with maximum flexibility, and extensibility that will permit new and existing applications to seamlessly extract data from an array of changing heterogeneous sources. The utility of many data bases in the IR field is seriously constrained by the difficulties of reconciling known disparities and conflicts within and across sources. (Data reconciliation itself has become an important focus of scholarly inquiry in various parts of International Relations).

Section 3. International Relations Research

3.1 Brief Domain Overview

The study of International Relations (IR) in Political Science generally converges around two seemingly distinct, but interrelated 'poles', namely matters of (a) conflict and war and (b) cooperation and collaboration. Both 'poles' address matters of sovereignty and security, national action and international consequence, local disruptions and global impacts, national integration and regional contestation – among others. Differences in theories, methods, and data practices create different perspectives on issues, shaping different "questions", and potentially leading to different "answers". The proliferation of new actors (i.e. states, non-governmental organizations, cross-border political groups, non-state actors, international institutions, global firms, etc.) reflects diverse perspectives, creates new sources of data and new difficulties for access, interpretation and management. Therefore, it comes as no surprise that fundamental changes in the international system have created new priorities and challenges for the conduct of research and the making of policy.

3.1.1 International Changes

The fall of the Berlin Wall, the end of Communism as a global threat, the demise of the Soviet Union, and the creation of new states with new configurations and strategic dilemmas are among the most significant and observable changes in the overall international context.. While there is a near-consensus about the salience of these changes, there is less agreement as to the nature, scale, and scope and, more importantly, the extent to which these alter prevailing patterns of 'politics as usual'. By the same token, new realities such as these have facilitated new venues for collaboration on a range of relatively 'new' issues, notably environmental degradation, electronic communication, regulatory strategies, etc.

It is not our purpose here to provide a review of the IR field and the underlying theoretical contentions, but rather to touch base with those aspects upon which we build our own research proposal, and to focus on the theoretical and empirical issues to which we expect to make some direct contributions. Our point of departure is reflected by a review of empirical challenges in a noteworthy issue of *International Political Science Review* (2001), devoted to "Transformation of International Relations – Between Change and Continuity". It argues that the "reconfiguration of the founding concepts of international relations … is linked to important paradigmatic changes" [Sind01, p. 224] and that state-centric modes of analysis and information configuration must be augmented by methods that help capture changes in both structure and process in the international arena.

3.1.2 Opportunity Cost

Under these circumstances, it is somewhat intriguing that the political science field as a whole has paid relatively little attention to the Internet, the changing scale and scope of information flows, and the forging of 'cyberspace', which has literally created a new domain of IR, known as 'cyberpolitics'. This is especially surprising given the strong and growing traditions of quantitative political analysis in many domains of political science. Interestingly, the *International Political Science Review* (2000) issue "CyberPolitics in International Relations" [Cho00] identifies new directions of research, research priorities, and critical next steps. But the profession's leading journal, the *American Political Science Review*, has yet to address these new domains, or to recognize attendant research challenges. This is of some irony, of course, since the United States is the world leader in

information technologies, and US political scientists continue to shape the field of IR.

While the provision of information through the Internet has become standard operating procedure in almost all endeavors, there are significant opportunity costs associated with barriers to the effective use of dispersed, diverse, and disconnected data sources. *Our goal is to reduce prevailing barriers, enhance understanding and meaning across substance, topics, and ontologies, and to provide new tools for IR research.*

3.1.3 Logic for Proposed Research

This goal is important because existing information systems are not easily comparable, nor do they readily interface. For example, there are data on incidences of conflict between nations located on the web sites of a wide range of institutions with different capabilities and objectives – such as the US Department of State, SIPRI (the Swedish institution focusing on peace research), the UN Higher Commission on Refugees, and the Correlates of War Project, to name a few. So, what is the 'real' incidence of conflict and the 'real' volume of casualties – at one point in time, over time, and as the contenders change and reconfigure their own jurisdictions? These are typical questions that have plagued researchers in the IR field, as far back as 1942, with classics in the field such as Quincy Wright's *A Study of War,* [Wri65] and even earlier, with Lewis Fry Richardson's *Statistics of Deadly Quarrels* (1917) [Rich60].

In order to (a) bound and define more precisely the proposed research strategy and (b) focus on its operational as well as analytical implications, we turn to the proposition at the onset of this proposal, namely that important research challenges are defined by the *new convergences* (i.e. globalization, world-wide connectivity, knowledge intensity) that shape *new information challenges* (i.e. information upsurge, new needs due to changes in content and contexts, etc.) noted in the opening section of this proposal. These challenges are evident across the two 'polar' aspects in the study of IR, namely *conflict and violence* and *cooperation and coordination*. We fully recognize that these two domains are extensive in scale and scope, differ in their theoretical underpinnings, and are not mutually exclusive in their content or coverage. Considerable advances in the field enable us to define specific gaps and needs that can be addressed rigorously by designing a cross-disciplinary and replicable research strategy. Here we focus largely on the conflict domain and concentrate on three modal types: (i) crises dynamics, (ii) conflict and war, and (iii) anticipation, preventative, and early warning. Most of the challenges we will address are also relevant to the collaboration domain, such as, modes of coordinated international action, approaches to peace-making, alignment of national and international responses (toward shared goals), and private sector cooperation (promoted by projects such as the Global Reporting Initiative (GRI)).

3.2 IR Research Needs

The proposed research strategy is framed by (1) *central tendencies* in the field and (2) *information gaps impeding theory development.*

3.2.1 Central Tendencies

While there exists no 'single authoritative view' of the field as a whole, Katzenstien, Keohane, and Krasner, eds. [KKK99], summarize two dominant perspectives in the field (labeled as rationalist and constructivist), both of relevance to conflict and cooperation. Their book is noteworthy for stressing differences as well as similarities across the two perspectives, but it is rather limited in its attention to quantitative data and information. For example, the chapter by Milner [Miln99], which assumes that states are the main actors in international relations, would have benefited from data on state formation and demise over time, comparisons with emergence of non-state actors, and a net assessment of the implications. In the absence of agreed upon metrics to track fundamental structural changes, IR theory remains dominated by assertions about, rather than, metrics of, change. In the segment of the field known as Quantitative International Politics (QIP), theory development is generally more data-driven and thus more vulnerable to the information limitations than other studies. Earlier QIP works, such as Hoole and Zinnes [HZ76] and Russett [Russ72], as well as the more recent advances by Levy [Levy89], Pollins and Schweller [SP99], and Choucri and North [ChoN93], illustrate the general progression in the field and the persistent data problems. Concurrently, [Alk96] highlighted some of the fundamental challenges to humanistic approaches to international studies, notably uses of computer-assisted applications.

In a related set of developments, some scholars in the field have given serious attention to interconnections between 'theory' and 'quantitative analysis' [Rose90]. Especially illustrative in this connection is the issue of *International Studies Quarterly* [CR96] devoted to evolutionary perspectives in international relations. Leading scholars such as George Modelski, Robert Gilpin, Cioffi-Revilla, and others, have begun to articulate the importance of transformation and adaptation over time, in contrast to the common focus on discrete events, or retrospective interpretation, which is quite dominant in the field. However, cumbersomeness in information access and data analysis makes it very difficult to replicate these works or to extend them in cumulative directions.

3.2.2 Impacts of Information Practices

There is a critical dilemma for researchers whose theoretical work relies on the use of systematic information and robust databases. Despite the *abundance* of existing data and information, there is a *paucity* in the consistency, reliability, and connectivity of the information. For example, in the conflict theory domain, the long tradition of tracking wars and casualties has been severely hampered by the difficulties of generating an integrated information system, drawing upon large scale efforts in the profession undertaken by a large number of different research groups. The same point holds for the cooperation theory domain where, for example, efforts to measure 'regime formation' and 'compliance' in a wide range of specific issue-areas are hampered by the diversity of ontologies, data meanings, and metrics. This dilemma common to both studies in IR is a *data and information disconnect* that appears at first glance to stem from data paucity, but is actually due to the inability to fully utilize the data compiled by different scholars on the same issue-areas.

Addressing the information disconnects will enable more intelligent access to existing databases and help to bridge the gap between conventional statistical analysis in the field and innovative modeling efforts to represent complexity in IR. For example, in the conflict domain, this will help us articulate and test propositions about potential linkages among long term pressures leading to antagonisms, the formation of escalation processes, the 'outbreak' of critical crises, and possibly the 'war' event as a distinctive outcome. In the cooperation domain, this would enable us to test for content and effectiveness of regimes by type, commitment, and potentials for durability, and across different issues, ranging from international treaties on environmental management to the non-proliferation of nuclear weapons.

3.3 Research Priorities

Our expectation is that advances in integrating critical pieces of information will allow the *whole* to yield insights and evidence greater than the sum of the *individual* parts. We seek to focus on applications of advanced information tools in order to reduce barriers to cumulativeness, and correct distortions due to data temporality

3.3.1 Reducing Barriers to Cumulativeness

In the social sciences, 'cumulativeness' refers to the extent to which advances in knowledge are based on previous findings, and the extent to which the linkages among them can be made explicit. We seek to understand exactly what findings derive from which theories and are shaped by what types of empirical data. Currently, advances in the field are difficult due to the lack of reliable ways to make appropriate inferences based on previous work. This difficulty stems from the differences in assumptions and theoretical perspectives and the inability to draw inferences *across* data sets about the *same* common phenomena. Efforts as those by Geller and Singer [GS97] in the domain of international conflict are laudable indeed, but have serious limitations in coverage and approach, leading to somewhat arbitrary conclusions due largely to selective review rather than comprehensive assessment of existing studies. Such limitations would be reduced substantially if there were greater ease of access to data and information from the very studies under review. Chronic difficulties in the field, such as these, seriously obstruct cross-method, cross-data, and cross-ontology comparisons. By the same token, one of the field's most innovative approaches to analysis of international conflicts (CASCON [BM97], see web.mit.edu/cascon/) is limited by constraints in cross-case comparisons and the difficulties of customizing information integration from multiple sources. Since CASCON is used in both the scholarly and the policy communities, reducing its current constraints will enhance its usefulness. In short, improving effective information access will increase propensities for cumulativeness in the field, in theoretical as well as empirical terms.

3.3.2 Correcting Distortions due to Data Temporality

We seek to understand the principles underlying initial compilation of data and their potential shifts over time, also referred to as the temporality of information. This is particularly relevant to certain issues within IR such as state integration and disintegration, alliance formation and dissolution, and cross-border activities and transnationalism. These issues blur the distinction between national politics and IR, and remind us that sources of insecurity can come from either of these domains, or both. For example, Walter and Snyder [WS99] point to critical features of local and civil wars that may generate international and global implications. This blurring of system boundaries between internal and external politics has important implications for information organization, management, analysis, and distribution; and these are likely to change over time. Given that the dominant practice has been to assume some form of unit stability (state boundaries, jurisdictions, etc.), it remains operationally very cumbersome to rescale or readjust observations given changes in boundaries, for example, and the attendant institutional responsibilities for national statistics. In addition, it is not uncommon for definitions of core terms to change, in response to changes in emerging 'realities', but our information practices continue to lag in this regard.

For example, changes in the meanings of terms such as 'citizens' (that defines the national population), taxes (that shape revenue sources), and boundaries (that determine jurisdictional responsibility) could potentially affect the way in which information is organized and the inferences that can be drawn.

3.3.3 Reminder of Operational Example

In Section 1 above, we pointed to an operational example – identifying select consequences of the war in the Balkans – and pointed to some of the necessary steps that must be undertaken in order to yield 'correct' answers to questions posed by different 'users'. Stylistic as it might seem, this example is fundamental as it highlights matters of changing boundaries, sovereignties, currencies, etc. that are critical to the very definition and determination of 'who gets what, when, and how' in the international domain. The research priorities defined above will shape our *research platform* and specific *research tasks* (with attendant goals and potential contributions).

3.4 Research Platform

As noted earlier, we frame our proposed work in the context of the GSSD knowledge-network. The *information base* for the GSSD 'laboratory' consists of web based resources from over 250 institutions worldwide, representing a diverse set of data by type, scale and scope that is then cross-referenced and cross-indexed for ease of retrieval and analysis, according to an integrated and coherent conceptual framework covering the knowledge domain. The domain consists of a hierarchical and nested representation spanning 14 key socio-economic 'sectors' of human activities, attendant known problems, scientific and technological responses, social and regulatory instruments, and modes of international collaboration and conflict resolution. GSSD is chosen as a research platform because it: (1) provides a *domain ontology* based on rigorous applications of social science theories, and related domains in science and technology, (2) offers practical reasoning rules for forming additional ontologies, (3) presents scenarios for broad applications of the integrated technologies to be developed in this project, (4) has identified a large and important set of information sources, and (5) spans local and global data sources.

3.5 Research Tasks and Expected Contributions

1. Undertake a comprehensive information-base survey. The goal of this task will be to fully understand attributes of the data types in the GSSD knowledge base that are relevant to international conflict. The anticipated contributions of this phase include: (a) **an assessment of** data types within the *conflict* domain, according to the following attributes: data source, format, organization, temporality attributes, provision rules, and utility for user-driven query; and (b) typologies of barriers to access (note sections 3.3.1 and 3.3.2 above).

2. Conduct an extensive multi-disciplinary and distributed user survey and develop test cases. The goal of this task will be to develop and apply methods to survey current and future information demands from diverse IR actors, differentiated in terms of (i) data users, (ii) data providers, and (iii) data intermediaries (or brokers). Test cases to capture the impacts of represent different user types on information and data needs will emerge from this assessment. The anticipated deliverables include: (a) **multi-dimensional assessments of information demand** from different user types within the diverse conflict domains noted earlier (e.g. sections 1.2.1 and 1.3), based on surveys, workshops, and in-depth interviews, and (b) a set of **IR test cases**, derived from the information demand and user surveys, illustrating information gaps, lags, and barriers and the opportunity cost inherent in applications of advances in IT to IR theories and methods.

3. Refine and develop ontologies and a knowledge repository to represent IR domains and provide a test bed for the emergent information technologies. The goal of of this task is to refine the GSSD ontology and develop select ontologies that are necessary to support the specific IR subject-domains, and to use this platform (and knowledge repository) as a testing ground for the proposed technologies. The anticipated contributions of this phase include: (a) **new and refined ontologies** related to the conflict/IR domain and (b) a **knowledge repository** to house the ontologies and information on applicable data sources available on the Internet.

4. Define the substantive features of the new technologies for enhancing information capabilities in IR theory and methods development, and test the effectiveness of the design. The goal of this task is to demonstrate the technologies' domain specific and practical applications of IR test cases and to explore relevance for similarly complex domains. The anticipated deliverables include: **collaborative assessments** of the technologies' effectiveness to address IR information issues and the architecture's capacity for scalability and cross-domain applicability, based on the following criteria: support for diverse information needs in a complex domain, as the salience, extent of customization, and complexity of the data demands vary, and robustness to changes in information properties and demands, given the diverse knowledge providers and the emerging global challenges and uncertainty in this increasingly complex world.

5. Enrich curriculum design and development. The goal of this task is to enrich educational development by incorporation and application of the information technologies developed in this project into (i)

information technology curricula, (ii) political science and other social science curricula, and (iii) multidisciplinary courses with strong international components. The anticipated deliverables of this phase include: (a) **on-line courses** on integration technologies and courses on conflict and war, drawing on the ontologies developed in this research and (b) with the help of our international collaborators, systematic tests of the **relevance of course design and implementation** in different regions of the world.

Section 4. Laboratory for Information Globalization and Harmonization Technologies and Studies

The **Laboratory for Information Globalization and Harmonization Technologies and Studies (LIGHTS)** will be established to address the strategy, application, development and deployment of intelligent information technologies that support the study of complex issues of the 21st century. Its purpose is to examine 'frontier' issues, such as transformations in patterns of conflict and cooperation, changes in modes of international business, emergent dimensions of globalization and system change, and negotiations for new global accords, among others. In addition to the research activities, the lab will host the technical infrastructure of the project, in particular our System for Harmonized Information Processing, and the publication and dissemination of research tools and findings.

In practice, the research activities in this multidisciplinary Laboratory will bring together faculty and students with interdisciplinary interests from a number of departments of MIT, including Information Technologies, Political Science, Management Science, and the Technology, Management and Policy program, as well as key research centers relevant to this work, notably the Center for eBusiness (CeB), Center for Technology, Innovation, and Policy Development (CTIPD), the Center for International Studies (CIS), and the Laboratory for Energy and Environment (LFEE).

The proposed Laboratory will be the central entity for producing products in four areas: (1) Software Platforms, (2) Knowledge Repositories, (3) Application Demonstrations, and (4) Education and Research. The software platforms will include but not be limited to: Collaborative Domains Space (CDS) Systems including one or more Ontology Library Systems, Context and Conversion Management Systems, Context Mediation Engine, Execution and Planning Module, and Application and Source Support Tools. The Knowledge Repositories will include both the structure and the content to define a significant portion of the knowledge needed for applications from International Relations. The IR domain specific knowledge will be represented in ontologies, context and conversion libraries, source schemas and capabilities. The Application Demonstrations will be developed at MIT, with the participation of the Project collaborators. There will be significant effort focused on technology transfer and open source Web presence. In Education and Research, the Laboratory will have three sets of outreach activities to the scholarly and the policy communities: (a) an ongoing Workshop on Innovations in Information Management, designed largely for experimental work across disciplines and domains, (b) a periodic Symposium on International Relations and Advances in Information Technology, targeted as an interface to the national and international policy-making communities, and (c) a web site that will include access to our System for Harmonized Information Processing, host the Studies, house ongoing research activities, and useful links that are relevant to our research, as well as electronic discussion forums. The Laboratory will also issue its own working papers and, as appropriate, organize its Book series, potentially with the MIT Press, and coordinate the Project's educational activities, research materials, and outreach initiatives.

Section 5. Educational Impacts

This multidisciplinary project addresses large-scale issues that will bring together graduate students with interdisciplinary interests from a number of departments of MIT. Integration of the research project into the education of these disciplines will train students to have multidisciplinary skills and prepare them for tackling even more complex problems in their research career.

We expect that the approach and technology platform developed in the project will be integrated in classrooms and be used for developing new curriculum, which will fundamentally change how knowledge is conveyed and significantly enhance the effectiveness of education. For example, political science students will be able to focus their effort on analyzing issues of crisis development and management without spending much time looking for, and reconciling relevant information; computer science students will be able to practice their skills by creating applications for other domains on top of the provided platform. We also plan to design new educational venues in "IR and IT" based on this proposed research to enable and facilitate multidisciplinary education and research. This initiative may take a number of forms, e.g., joint supervision of Ph.D. students; hosting post-doctoral researchers; knowledge dissemination and experience sharing through seminar series and regular workshops, etc.

We anticipate that the impact to education will be profound and continuous as our international collaborators begin to adapt the project's curricula to their own contexts, educational programs, and institutional conditions.

Section 6. Anticipated Contributions

The project will lead to major advances in information technology and revolutionary approaches to international relations research and related fields. The outcomes of this innovative project will address head on many of the challenges in each of the four NSF ITR multidisciplinary focus areas; some examples, noted in bold, include:

1. Software and Hardware Systems. This project will enable us to create a **robust platform,** the LIGHTS System for Harmonization of Information Processing (SHIP), for **meaningful information interchange** among **very large scale** (in terms of size and geographical locations) and **diversified** (in terms of media, schemas, and domains) systems. **Reliability** of systems built on this platform will be significantly improved by dynamically incorporating semantically equivalent sources into the interconnected system. The general-purpose platform will allow new applications to be built quickly to **facilitate information sharing among diverse groups of people, devices**, and software systems. Since the platform will facilitate semantic level information interchange, any information receiver (people, devices, or software) can get information accurately and in a form and meaning that the receiver prefers.

2. Augmenting Individuals and Transforming Society. This project will significantly **augment the effective use of information in our society** and **expand the frontiers of political science and information technology**. This has important applicability for **increasing national security** and **prevention and attribution of terrorism**. We intend to generate empirically-based and systematic insights into how people access and use **large-scale heterogeneous data sources** in a complex domain like IR. These findings will help us to define the requirements for the necessary Collaborative Domain Spaces (CDSs) and meet the goal of **improved information utilization** that also can be applied and extended to other complex fields of study. Through **international collaborators** we will be able to obtain a more robust handle on matters of context, culture, multiple interpretations, multilingualism, imperatives of localization, etc. that will invariably continue to shape the nature of **international relations**. Our approach will advance the frontiers of political science by providing a powerful tool for information-intensive analytical frameworks, which will **change the conduct of political science research**. This contribution also will lead to **more effective use of information in society** enabling **more informed citizen participation**.

3. Scientific Frontiers and Information Technology. A key product of this research will be the operational capability for effective **domain and context knowledge discovery**. The SHIP will enable us to gather data from **large-scale heterogeneous sources** and intelligently and effectively **interpret and integrate** it – making possible **the creation of consistent data sets over vast scales of space and time, as well as generating new data sets.** This will enable strategic decisions to be made timely and informatively.

4. Knowledge acquisition and interpretation. Two of the fundamental goals of this project are (1) the **acquisition** of information **context knowledge** (both for sources and users) and (2) the ability to use our proposed SHIP's reasoning ability about this knowledge to correctly and effectively **organize and interpret** the information.

5. Education. Our project will contribute to education in many ways: it will help to transform the traditional IT educational setting by incorporating various disciplines into the development of new IT theories and tools. Similarly, political science students (and related social sciences) will advance their understanding of complex issues in their field through the use of these technologies, and advance the field by focusing on analysis rather than on the diversion of reconciling disparate data. In addition, by facilitating the **integrated study** of complex issues, this research will help to develop and foster new multidisciplinary learning environments. Our project will also contribute to the education of new researchers, including post-doctoral associates, graduate students, and undergraduate students, who will take an active role in the research of this project. We propose to interface with the MIT OpenCourseWare administration to draw on the most recent educational technology outreach system.

In conclusion, the research team plans to utilize the Internet and the technical infrastructure developed by the new Laboratory for Information Globalization and Harmonization Technologies and Studies (LIGHTS) to share its findings and encourage collaboration with the broader research community**.** The materials that will be publicly available on the Internet include: literature reviews, survey results, theoretical models, reports, the System for Harmonized Information Processing technology, other analyses conducted during the life cycle of the project, and a discussion forum. This approach serves three purposes: potential materials of interest are provided to the intellectual community in a more timely manner than would be possible with traditional academic publications; the range, scale and scope of outreach are considerably expanded; and the potential for timely and valuable feedback on the research is significantly enhanced. We expect the results will generate profound impacts for the research, education, and various practitioner communities, as well as society, in general.

PROJECT MANAGEMENT PLAN

Recognizing that advances in information technology are essential for achieving the Nation's 21 century aspirations, we propose to integrate and manage all components of the proposed research under a newly created laboratory, named the **Laboratory for Information Globalization and Harmonization Technologies and Studies (LIGHTS).** The lab will oversee all research activities, host the technical infrastructure, coordinate outreach activities of the project, and disseminate the products of LIGHTS research (such as publications, platforms, tools, and educational materials) and host the proposed Symposia and Workshops.

The laboratory will be jointly run by the co-PIs (Choucri, Madnick, Siegel, Wang) who have effectively worked together (in groups of two or three) on other projects. One of the PIs (Siegel) will take the key role in the day-to-day management and coordination of the Laboratory. This management team is dedicated to providing results that will directly address the information technology problems and applications central to national priorities in IT.

A **steering committee** of approximately eight individuals will be formed from the national and international collaborators, drawing approximately one individual from each of the categories listed below. This steering committee will meet at least twice annually and provide both feedback and priorities to this research effort.

The proposed project is composed of three components that will focus on different, but related, areas of interest: (1) identifying barriers to access of information for education, research, decision making, and performance in the complex domain of international relations (**IR**), (2) development of new information technologies (**IT**) to address these needs for both IR and similarly complex domains, where there are multiple actors and domains of salience, and rapidly changing conditions, and (3) advancing developments in the use of the technologies to facilitate interdisciplinary research and contribute to new education materials, approaches, tools, and methods.

The IR research component will be directed by one PI (Choucri) and will include the efforts of one full-time doctoral student and several research assistants. The IT development will be directed by one PI (Madnick), with specific technical areas assigned to the other co-PIs (Siegel and Wang), and will include the efforts of one full-time doctoral student and several research assistants. The education component of the project will be supported by all four PIs, and will include the efforts of all full-time doctoral students and graduate research assistants. All of the PIs have considerable prior experience with the organization and management of large scale, international, distributed, and diverse research projects.

At the foundation of this proposal is a network-in-place of national and international collaboration. These include a wide range of collaborators, each with their own distinctive operational context and expected participation. The list below names some of the initial collaborators that have verbally committed to this effort (*letters of confirmation from fourteen of the collaborators, marked with *, have been received in time to be included in the Supplemental Documents*). The Table highlights four types of contributions: (1) **reviewers** (who contribute valuable input on the research), (2) **data sources** (who provide data for application testing), (3) **users** (potential users of the technology who help with the problem definition and who provide challenging test cases), and (4) **active researchers** in either **IR** and/or **IT** (who will directly participate in and contribute to our research). None of these collaborators will be receiving any of the NSF funds, but they will significantly leverage the funds that are provided.

Names and Institutions of Collaborators	Institution Type	Anticipated Roles	Benefits to the Research
C. von Furstenberg, **UNESCO** B. Plescovic, **World Bank**	International governmental organizations	**Data sources and users,** contribute to understanding changing policy contexts & impact on information needs.	Direct inputs on policy deliberations affecting context and framework for of international information systems.
J. Cares, **Alidade Consulting** * M. Laguerre, **U. Berkeley Institute of Global Studies** * P. Brecke, **Georgia Tech, Nunn School of International Affairs** B. Pollins, **Ohio State University** M. Feldman , **Stanford University** A. White and R. Massie, **Global Reporting Initiative**	Scientific research and policy institutions	**Reviewers, users, and active researchers (IR),** who will also participate in workshops and help to develop new applications.	Provide comparative bases for assessing generalizability and collaborate on new applications.

Names and Institutions of Collaborators	Institution Type	Anticipated Roles	Benefits to the Research
* B. Allenby, **AT&T** * W. R. Baker,**Baker & McKenzie** * Dan Schutzer, **Citibank** U. Wennberg, **Global Responsibility, International** K. Cavanaugh, **IBM** * J. D. Funk, **S.C. Johnson Company** * L.G. Scheidt, **Sony International Advanced Technology Center**	Global firms – Information Technology, Legal Services, Financial Services, Consumer Products, and Electronics	**Reviewers and users,** contributing to improved applications, including relevance of changing contexts. Insights into integration issues in large multinational environments with heterogeneous global data sources.	Diversity of professional and domain expertise, covering variations in legal contexts, environmental research, and responses to the cultural diversification of the global workplace. These organizations are currently working with various of the co-PIs.
* B. Davidson, **Cedars Sinai Health System** * C. Marshall, **New York State Office for the Aging**	Non-profit org – health care and elderly	**Reviewers and users,** important applications and issues in complex governmental and non-profit environments with heterogeneous data sources.	Currently working with co-PI Wang on improving the use of information in their organizations, especially improving information quality.
G. Kochendoerfer-Lucius, **German Foundation for International Development** C. Brodhag, **Ecole des Mines a St. Etienne**, **France** S. Chengyoung, **Ministry of Science & Technology**, **China**	Governmental scientific agencies	**Data source and active researchers (IR/IT),** contributing to contextual evaluation, cross-cultural interpretation and meanings, local knowledge provision, and comparison across contexts.	Currently working with PI Choucri on global knowledge networking. Direct input into contextual biases, or errors in assignment of meaning to recorded observations.
* T. Mezher, **American University of Beirut, Lebanon** A.Koshla, **Development Alternatives India** M. Tolba, L. Hassenien, **ArabDev, Egypt**	Researchers from institutions in developing countries	**Data source and active researchers (IR),** with a focus on the provision of local and national knowledge.	Currently collaborating with PI Choucri on global knowledge networking. Important to comparative and diverse contextual applications, validation of internationalization.
* F. Manola, **MITRE**	Non-profit corp. operating federally funded R&D centers	**Data sources, reviewers, users, active researchers (IT),** providing input in theory and system development. Liaison with federal agencies	Currently working with Co-PIs Madnick and Siegel on issues related to Semantic Web.. Substantial knowledge representation and technology experience.
* A. Segev, **U. Berkeley Center for information Technology** * Nor Adnan Yahaya, **Malaysia University of Science and Technology** * Tan Kian Lee and Stephane Bresson, **National University of Singapore**	Research Universities	**Data sources, reviewers users, active researchers (IT)**, providing complementary labs for development of theory and software platform.	Working with Co-PIs Madnick and Siegel. Active database researchers having significant experience with web-based information integration.

Due to the highly multi-disciplinary nature of this effort, the research will be supported by this outstanding and diverse research team of international collaborators, with multiple demographics, experiences, and qualifications. We believe that this project will lead to important developments in the areas of IT and political science/IR. In particular, their intersection will have a significant impact on the way organizations (e.g., governments, companies, world bodies) understand, react to, and manage the significant global challenges (e.g., war, terrorism, environment) of the 21st century.

FACILITIES, EQUIPMENT, AND OTHER RESOURCES

COMPUTING EQUIPMENT AND DATA SOURCES

The **Laboratory for Information Globalization and Harmonization Technologies and Studies**, to be formed, will primarily use existing computing equipment from the Context Interchange Systems (COIN) laboratory (within the Information Technology group of MIT's Sloan School of Management) and the Global System for Sustainable Development (GSSD) project (within MIT's Center for International Studies and Political Science department.) Both facilities are located in the same building, and most on the same floor, so coordination will be easy.

Equipment currently available within the COIN lab includes two Sun Unix servers, two Windows NT servers, a Linux server, and 16 current generation Intel workstations running versions of Windows or Linux as appropriate for research needs. Available software includes Microsoft development, systems, and server platforms as well as open source resources for software development, knowledge management, and database management. The latest version of the COIN context mediation prototype, for knowledge representation and reasoning, was developed within this lab and this software infrastructure will constitute a starting point for the proposed effort.

In addition, we will draw on the two Pentium workstations and 3 Windows NT servers, and data sources of the GSSD. GSSD is the knowledge networking and management system for the Alliance for Global Sustainability (which includes MIT, University of Tokyo, Chalmers University-Sweden, and ETH - the Swiss Technical University System). GSSD mirror sites are maintained in France (École Nationale Supérieure des Mines de Saint Etienne), China (Ministry of Science and Technology) and Japan (University of Tokyo).

OTHER RESOURCES

As part of its dual and integrated focus on education and research, there are more than 3,000 ongoing projects on campus at MIT. These projects utilize shared centralized facilities, such as contemporary computational aids and library facilities, as well as specialized facilities of individual departments, research centers, and labs. Each project is affiliated with a nodal department, but can access resources in other parts of MIT. This project will draw particularly on MIT's extensive communications and network infrastructure.

The co-PIs are affiliated with various organizational units and research centers at MIT and will have access to their resources, especially the departments of Information Technologies, Political Science, Management Science, and the Technology, Management and Policy program, as well as key research centers, notably the Center for eBusiness (CeB), Center for Technology, Innovation, and Policy Development (CTIPD), the Center for International Studies (CIS), the Technology and Development Program (TDP), the Total Data Quality Management (TDQM) program, the Productivity from Information Technology (PROFIT) program, and the Laboratory for Energy and Environment (LFEE).

A primary mission of MIT is education and many MIT resources will be used to facilitate the development, testing, and deliver of new educational materials. In particular, we plan to work with MIT's OpenCourseWare initiative, which reflects MIT's institutional commitment to disseminate knowledge across the globe. One of the co-PIs (Madnick) serves on the OCW Advisory Board. We will also make use of other media development, presentation, and transmission facilities, such as MIT's new Learning International Network Consortium (LINC), which supports collaboration and cooperation across international borders through technology-enabled media for higher quality education to 'learners' worldwide.

In this research effort we plan to work with collaborators as reviewers, data sources (who provide data for application testing), users (potential users of the technology who help with the problem definition and who provide challenging test cases), and active researchers in either IR and/or IT (See Management Plan for more details). As a result of the active participation of these collaborators (i.e., international and governmental organizations, scientific research and policy institutions, researchers from institutions in developed and developing countries, global commercial firms, non-profit organizations and universities) we expect to have access to and involve a number of resources from these organizations, including databases, applications, algorithm and theory development, software, and facilities for meetings and demonstrations.

BUDGET JUSTIFICATION

Key Personnel

This project represents a major multidisciplinary effort with significant distinct but interrelated components: (1) Information Technology development, (2) applications and studies in the complex domain of International Relations, (3) knowledge collection, (4) educational material development, and (5) outreach for education and global impact. We expect the Principals to lead these efforts, to coordinate across the components, and to facilitate their success. A post-doctoral student will work alongside Dr. Siegel to facilitate coordination across these diverse efforts, between the schools at MIT, and with our national and international collaborators. We recognize the importance of ensuring timely activities and outputs, appropriate sequencing of tasks, and effective streamlining of interactions among all participants, as well as managing report preparations, working papers, and internal and external research communication. We feel that these efforts will require the times allocated by the PIs and the post-doc.

This is a substantial project with global scope that will require careful support, in terms of clerical requirements as well as financial tracking and analysis. For this reason, we plan to fund 1/3 of the time of an administrative assistant. We believe this to be very conservative given our overall needs. In addition, for efforts such as conferences, courses, and larger meetings we can call upon our undergraduates and graduate students to assist. Such involvement has a multiplier benefit, namely that of providing these students with a closer working relationship with the project, tighter connections to their educational programs, and new experiences working with its collaborators -- while limiting our support staff requirements. In addition we allocated 1/10 of a person time to assist in management of budget, internal MIT requirements, and financial reports required of a large research project

This project seeks to, and will, depend heavily on graduate and undergraduate students, as an important contribution to their education (in terms of basic research as well as the 'pre-testing' of educational materials we will prepare). For graduate students we will be using one doctoral and two masters students to focus on the development of the technology platform, SHIP. For the collection of IR related data, the generation of new data, and integration of information on key data-generating institutions, we will be using one doctoral and one masters student. All students will be assisting in the development of coursework, meetings, seminars and other outreach programs. Graduate students are noted in Section B of the budget. Section B.6 is for summer support of graduate students (noted as Other).

We intend to involve undergraduate students as well, to assist the project in select tasks. More specifically, these students will be UROP (Undergraduate Research Opportunity Program) students. UROP is a unique program at MIT that has a long track record of providing undergraduates the ability to work with research projects while providing faculty with a low cost, low overhead, high motivation workforce. These students have proven to be particularly helpful in developing software, collecting data, and supporting faculty and advanced graduate students.

MIT budgeting guidelines:

(1) Non-faculty salaries are inflated at 3% per year on January 1 of each year
(2) Faculty salaries are inflated at 3% per year on July 1 of each year

Equipment

This category includes all equipment with cost over $3000. At these costs MIT does not charge overhead on such equipment. Also, it is possible to make upgrades to this equipment without occurring additional overhead. A majority of the machine costs are included in this category because of the favorable overhead situation. Complete systems including peripherals can be priced in a bundle to fit into this budget line.

In the first year we plan to use a larger sum ($15,400) to set-up two servers and 3 complete desktops. The servers will be used to support a development platform and a release platform to collaborators and other organizations, nationally and internationally. It will support the website for outreach and education programs and the center point for dissemination of publications. Access to all applications, domain knowledge and contributions by collaborators will be coordinated through this site. The three desktops will be used to support the post-doc and two of the graduate students.

After the first year we will use portions of the smaller equipment budget ($5400 in year 2 and 3) to upgrade existing machines and one new machine per year to replace older machines that were in place prior to the project but that have become obsolete. In the last year we use a smaller budget ($1600) to upgrade existing machines.

A smaller budget ($3200 first year and $1800 in later years) is used for smaller equipment (fax machines, printers) and networking equipment and one desktop per year for students or principals. Overhead is applied to such equipment purchases.

Travel

Travel is central to our outreach activities. In all years we plan one trip for two principals for meetings with NSF and other government agency interested in the research and the results of the research. We target four trips to meet with collaborators, these would be both domestic and international. These would be meetings where we can get feedback on our application development, access to new sources of information, additional technology and domain expertise, and transfer results. In addition we plan to submit and present publications at both domestic and international conferences and have budgeted a total of three conference attendances to allocate over our faculty and students at a cost of ($6000 per year total).

Our travel is important for outreach. However to accommodate the number of faculty and students we plan to use MIT as the primary meeting place for collaboration. Meeting costs will be discussed in the section on Other (G.6)

Materials and Supplies

These are based on a first year rate of $6100. These include all costs for postage, telecommunications, network communication charges, back-up charges, and office supplies. These will be used by the individual investigators and for the new laboratory.

Publications/Documentation/Dissemination

We plan to run a number of meetings/workshops and develop courseware for MIT's Open Courseware Initiative. Publication and dissemination costs are escalated in Years 2, 3, and 4 (i.e., by $2000) to accommodate a significant level of publication and outreach. Courses will be developed based on this research for the Sloan School of Management, the Political Science Department and for use in MIT-Singapore and MIT-Malaysia Alliances.

Computer Services

This includes a reasonable number for software purchases such as licenses for the Laboratory and individual licenses as needed.

Other (G.6)

This category ($60,000 first year) includes tuition expense for our students ($55000 first year no overhead charged), expenses for initial meetings with our collaborators ($5000 first year). Tuition costs are similar in all four years. However meeting costs vary. In the second year (10K meeting expenses) and we are planning a larger symposium with collaborators and outside organizations to present results and gain further buy-in to the project from the broader IR community across research, business, and policy domains and at regional as well as global levels. We expect to capture both the multiple perspectives and the diversity of organizational and institutional impacts upon information generation, compilation, and dissemination worldwide. In the third year we plan two smaller meetings/workshops at MIT with a larger group of collaborators. These will be working meetings and launch of application to selected collaborators. Finally in Year 4 we will host a broad workshop and application rollout to the community and transfer technology to appropriate collaborators to support a global platform for IR disciplines. Expenses for these meetings include costs for space, audiovisual equipment, networking, refreshments, and appropriate office supplies. These meetings and workshops will be web cast to reach the largest possible audience and minimize travel expenses.

Leveraging of these budgeted resources

The requested budget resources will be significantly leveraged in several ways: (1) much of the critical initial basic research, especially for the COIN and GSSD platforms, have been previously funded from NSF, DARPA, and industry sources, (2) the LIGHTS collaborators are committing significant internal resources that will greatly assist this effort as well to facilitate the dissemination of results and impact of this research, (3) relevant and unique resources at MIT, such as MIT's OpenCourseWare and LINC technologies and distribution resources, will be utilized; as well as (4) recent experience and ongoing participation in national and international policy-related exercises – such as the National Academy of Sciences (NAS) Committee initiative on Terrorism, DARPA-National Academy of Sciences Committee on *Understanding Terrorism in order to Deter Terrorism,* and the United Nations – Information and Communication Taskforce (UN-ICT) background work in preparations for World Summit on Information Technology (WSIS).

CITED REFERENCES

[AKS96] Arens, Y., Knoblock, C. and Shen, W.M. (1996). "Query Reformulation for Dynamic Information Integration." *Journal of Intelligent Information Systems*, 6: 99-130.

[Alk96] Alker, H. R. (1996). Rediscoveries and Reformulations: Humanistic Methodologies for International Studies. Cambridge, UK, Cambridge University Press.

[BGL*00] Bressan, S., Goh, C., Levina, N., Madnick, S., Shah, S., Siegel, S. (2000) "Context Knowledge Representation and Reasoning in the Context Interchange System", *Applied Intelligence: The International Journal of Artificial Intelligence, Neutral Networks, and Complex Problem-Solving Technologies,* **12**(2): 165-179.

[BM97] Bloomfield, L.P., Moulton, A. (1997) Managing International Conflict: From Theory to Policy. New York: St. Martin's Press. (supplemented by CASCON web site http://mit.edu/cascon).

[CFM*01] Chen, X., Funk, J., Madnick, S., Wang, R. (2001) "Corporate Household Data: Research Directions", *Proceedings of the Americans Conference on Information Systems*, AMCIS'01, Boston, August.

[Chen99] Chen, P (1999). "ER Model, XML and the Web". *18th International Conference on Conceptual Modeling*.

[CZ98] Cherniak, M. and Zdonik, S (1998). "Inferring Function Semantics to Optimize Queries". *VLDB*, 239-250.

[Cho00] Choucri, N. (2000). "CyberPolitics in International Relations." International Political Science Review. **21**(3): 243-264.

[ChoN93] Choucri, N., North, R.C. (1993). "Growth, Development, and Environmental Sustainablity: Profiles and Paradox." Global Accord: Environmental Challenges and International Responses. Cambridge, MA, MIT Press. 67-131.

[Chu00] Chu, W. (2000): "Introduction: Conceptual Models for Intelligent Information Systems". *Applied Intelligence 13*, (2).

[CR96] Cioffi-Revilla, C. (1996). Origins and Evolution of War and Politics. International Studies Quarterly. Hart, Rasler, Thompson, eds. Cambridge, MA, Blackwell Publishers, Inc. **40:** 1-22.

[DFen01] Ding, Y., Fensel, D. (2001) "Ontology Library Systems: The Key to Successful Ontology Reuse", International Semantic Web Working Symposium, Stanford University, CA.

[DSW*99] Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B., Bejamins, V.R. (1999) "WonderTools? A Comparative Study of Ontological Engineering Tools", Proceedings of the 12th International Workshop on Knowledge Axuisition, Modeling, and Management (KAW'99), Banff, Canada.

[DT95] Dobbie, G. and Topor, R. (1995). "On the declarative and procedural semantics of deductive object-oriented systems," *Journal of Intelligent Information Systems*, 4: 193-219.

[Fensel01] Fensel D. (2001) Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, Springer.

[FGM02] Firat, A., Grosof, B., Madnick, S. (2002) "Financial Information Integration In the Presence of Equational Ontological Conflicts," *Proceedings of the Workshop on Information Technology and Systems*, Barcelona, Spain, December 14-15: 211-216

[FMS00a] Firat, A., Madnick, S., Siegel, S. (2000) "The Caméléon Web Wrapper Engine", *Proceedings of the VLDB2000 Workshop on Technologies for E-Services,* September 14-15.

[FMS00b] Firat, A., Madnick, S., Siegel, S. (2000) "The Caméléon Approach to the Interoperability of Web Sources and Traditional Relational Databases," *Proceedings of the Workshop on Information Technology and Systems*, December.

[Fynn97] Fynn, K.D. (1997) *A Planner/Optimizer/Executioner for Context Mediated Queries*, MS Thesis, MIT.

[Gams68] Gamson, W. A. (1968). Power and discontent. Homewood, Illinois, The Dorsey Press.

[GBM*99] Goh, C.H., Bressan, S., Madnick, S., Siegel, S. (1999) "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information", *ACM Transactions on Information Systems*, 17(3): 270-293.

[GoBM96] Goh, C.H., Bressan, S., Madnick, S.E., and Siegel, M.D. (1996). "Context Interchange: Representing and reasoning about data semantics in heterogeneous systems," *Sloan School Working Paper #3928*, Sloan School of Management, MIT, 50 Memorial Drive, Cambridge MA 02139.

[Goh96] Goh, C. (1996). *Representing and Reasoning about Semantic Conflicts* In *Heterogeneous Information System*, PhD Thesis, MIT.

[GS97] Geller, D.S., Singer, J.D. (1997). Nations at War: A Scientific Study of International Conflict. Cambridge, England, New York, NY, Cambridge University Press.

[HelfH00] Helfin, J. and Hendler J. (2000) "Dynamic Ontologies on the Web", Proceedings of the 11th National Conference on Artificial Intelligence (AAAI-2000), Menlo Park, CA: 443-449.

[HZ76] Hoole, F.W., Zinnes, D. A. eds. (1976). Quantitative International Politics: an Appraisal. New York, NY, Praeger.

[KKK99] Katzenstein, P., Keohane R., and Krasner, S. eds. (1999). Exploration and Contestation in the Study of World Politics. Cambridge, Massachusetts, The MIT Press.

[KKT93] Kakas, A.C., Kowalski, R.A., and Toni, F. (1993). "Abductive logic programming," *Journal of Logic and Computation*, 2(6):719--770.

[KLW95] Kifer, M., Lausen, G., and Wu, J. (1995). "Logical foundations of object-oriented and frame-based languages," *JACM*, 4:741--843.

[Las58] Lasswell, H.D. (1958). Politics: Who Gets What, When and How. New York, NY, McGraw-Hill.

[LCN*99] Lee, T., Chams, M., Nado, R., Madnick, S., Siegel, M. (1999) "Information Integration with Attribution Support for Corporate Profiles", *Proceedings of the International Conference on Information and Knowledge Management,* November: 423-429.

[Levy89] Levy, J. S. (1989). The Causes of War: A Review of Theories and Evidence. Behavior, Society, and Nuclear War. Tetlock, Jervis, Stern, and Tilly, eds. New York, Oxford University Press.

[LMB98] Lee, T., Madnick, S., and Bressan, S. (1998) "Source Attribution for Querying Against Semi-Structured Documents", *Proceedings of the ACM Workshop on Web Information and Data Management* (WIDM'98), Washington, DC, November 6: 33-39.

[LMS96b] Lee, J., Madnick, S., Siegel, M. (1996) "Conceptualizing Semantic Interoperability: A Perspective from the Knowledge Level", *International Journal of Cooperative Information Systems*: [Special Issue on Formal Methods in Cooperative Information Systems], 5(4), December.

[Mad01] Madnick, S. (2001) "The Misguided Silver Bullet: What XML will and will NOT do to help Information Integration," Proceedings of the *Third International Conference on Information Integration and Web-based Applications and Services*, IIWAS2001, Linz, Austria, September: 61-72.

[Mad99] Madnick, S. (1999) "Metadata Jones and the Tower of Babel: The Challenge of Large-Scale Heterogeneity," *Proceedings of the IEEE Meta-data Conference*, April.

[MBM*98] Moulton, A., Bressan, S., Madnick, S., Siegel, M. (1998) "Using an Active Conceptual Model for Mediating Analytic Information Exchange in the Fixed Income Securities Industry", *Proceedings of the 17th International Conference on Conceptual Modeling* (ER'98), Singapore, November.

[McN00] McNeill, J.R. (2000). Something New Under the Sun. New York, NY, W.W. Norton & Company.

[MHR00] March, S., Hevner, A., Ram, S. (2000) "Research Commentary: An Agenda for Information Technology Research in Heterogeneous and Distributed Environments", Information Systems Research, **11**(4): 327-341.

[Miln99] Milner, H.V. (1999). Rationalizing Politics: The Emerging Synthesis of International, American, and Comparative Politics. Katzenstein, P., Keohane R., and Krasner, S. eds. Exploration and Contestation in the Study of World Politics. Cambridge, MA, The MIT Press. 119-146.

[Nuna01] Nunamaker, J. F. (2001): "Collaboration Systems and Technology Track – Introduction". *Hawaii International Conference on System Sciences*.

[RAC01] Richard, D.J., Allenby, B.R., and Compton, W.D., eds. (2001). *Information Systems and the Environment*. National Academy of Engineering. Washington, D.C., National Academy Press.

[Rich60] Richardson, L.F. (1960). Statistics of Deadly Quarrels. Pittsburgh, PA, Boxwood Press.

[Rose90] Rosenau, J.N. (1990). Turbulence in World Politics: A Theory of Change and Continuity. Princeton, NJ. Princeton University Press.

[Russ72] Russett, B. M. ed. (1972). Peace, War, and Numbers. Beverly Hills, Sage Publications.

[Sind01] Sindjoun, L. (2001). "Transformation of International Relations – Between Change and Continuity." International Political Science Review **22**(3): 219-228

[SM91a] Siegel, M. and Madnick, S. (1991). "Context Interchange: Sharing the Meaning of Data," *SIGMOD RECORD*, **20**(4), December: 77-8.

[SP99] Schweller, R. and Pollins B. (1999). "Linking the Levels: The Long Wave and Shifts in U.S. Foreign Policy, 1790-1993." American Journal of Political Science **43**(2): 431-464.

[SW92] Sheng, O. R. L., Wei (1992), C-P, "Object-Oriented Modeling and Design of Coupled Knowledge-base/Database Systems". *International Conference on Data Engineering*: 98-105.

[TBD87] Templeton, M., Brill, D., Dao, S.K., Lund, E., Ward, P., Chen, A. L.P., and MacGregor, R. (1987). "Mermaid --- a front end to distributed heterogeneous databases," *Proc of the IEEE*, **75**(5): 695-708.

[TM98] Tu, S.Y., Madnick S. (1998) "Incorporating Generalized Quantifiers into Description Logic for Representing Data Source Contents", *Data Mining and Reverse Engineering: Searching for Semantics,* Chapman & Hall.
[Van99] Van Evera, S. (1999). Causes of war : Power and the roots of conflict. Ithaca, Cornell University Press.
[WKM93] Wang, R., Kon, H., and Madnick, S (1993). "Data Quality Requirements Analysis and Modeling", *International Conference on Data Engineering*, 670-677
[Wied92] Wiederhold, G. (1992). "Mediation in the Architecture of Future Information Systems", IEEE *Computer*, **25**(3): 38-49.
[Wied99] Wiederhold, G. (1999). "Mediation to Deal with heterogeneous Data Sources", Proceedings of Interop'99, Zurich, March: 1-16.
[Wri65] Wright, Q. (1965). A Study of War. Chicago, IL, University of Chicago Press.
[WS99] Walter, B. and Snyder, J (1999). Civil Wars, Insecurity, and Intervention. New York, Columbia University Press

OTHER RELEVANT REFERENCES

Information Technology (IT)

[ACH93] Arens, Y., Chee, C., Hsu, C., and Knoblock, C. (1993). "Retrieving and integrating data from multiple information sources," *International Journal on Intelligent and Cooperative Information Systems*, http://www.isi.edu/sims/papers/93-sims-ijicis.ps.
[ASD91] Ahmed, R., Smedt, P.D., Du, W., Kent, W., Ketabchi, M.A., Litwin, W.A., Raffi, A., and Shan, M.C. (1991). "The Pegasus heterogeneous multidatabase system," *IEEE Computer*, 24(12): 19-27.
[BFG*97] Bressan, S., Fynn, K., Goh, C., Madnick, S., Pena, T. and Siegel, M. (1997) "Overview of a Prolog Implementation of the COntext INterchange Mediator", *Proceedings of the Fifth International Conference and Exhibition on the Practical Applications of Prolog*, London, England, April 1997.
[BFG96b] Bressan, S., Fynn, K., Goh, C.H., Madnick, S., Pena, T., and Siegel, M. (1996). "Overview of a prolog implementation of the COntext INterchange mediator," In *Practical Applications of Prolog 97*.
[BGF*97a] Bressan, S., Goh, C., Fynn, K., Jakobisiak, M., Hussein, K., Kon, H., Lee, T., Madnick, S., Pena, T., Qu, J., Shum, A. and Siegel, M. (1997) "The Context Mediator Prototype" *Proceedings of the ACM Conference on the Management of Data (SIGMOD '97*), February.
[BGT*97] Bressan, S., Goh, C.H. , Lee, T., Madnick, S., Siegel, S. (1997) "A Procedure for Mediation of Queries to Sources in Disparate Context", *Proceedings of the International Logic Programming Symposium,* October.
[CHS91a] Collett, C., Huhns, M., and Shen, W. (1991). "Resource Integration Using an Existing Large Knowledge Base," *MCC Technical Report Number ACT-OODS-127-91*.
[CHS91b] Collet, C., Huhns, M.N., and Shen, W.M. (1991). "Resource integration using a large knowledge base in Carnot," *IEEE Computer*, 24(12): 55-63.
[Date87] Date, C. (1987). *An Introduction to Database Systems, volume 1, fourth edition*. Addison-Wesley Publishing, Menlo Park, CA.
[DG96] Duschka, O. and Genesereth, M. (1996). "Query Planning in Infomaster," http://logic.stanford.edu/people/duschka/papers/Infomaster.ps.
[DGH*96] Darawala, A., Goh, C., Hofmeister, S., Madnick, S. and Siegel, M. (1996) "Context Interchange Network Prototype" in *Database Applications Semantics*, Chapman & Hall (London, UK), October.
[DGH95] Daruwala, A., Goh, C.H., Hofmeister, S., Hussein, K., Madnick, S., and Siegel, M. (1995). "The context interchange network prototype," In *Proc of the IFIP WG2.6 Sixth Working Conference on Database Semantics (DS-6)*, Atlanta, GA.
[FDF95] Faquhar, A., Dappert, A., Fikes, R., and Pratt, W. (1995). "Integrating information sources using context logic," In *AAAI-95 Spring Symposium on Information Gathering from Distributed Heterogeneous Environments*.
[FLM*01a] Fan, W., Lu, H., Madnick, S., Cheung, D. (2001) "Discovering and Reconciling Value Conflicts for Numerical Data Integration," *Information Systems Journal*, **26**(8), November: 635-656.
[FLM*01b] Fan, W., Lu, H., Madnick, S., Cheung, D.W. (2001) "DIRECT: A System for Mining Data Value Conversion Rules from Disparate Sources, *Decision Support Systems* (in press).
[GGKS95] Geddis, D., Genesereth, M., Keller, A., and Singh, N. (1995). "Infomaster: A Virtual Information

System," http://logic.stanford.edu/papers/iiaw95_infomaster.ps.

[GMP95] Garcia-Molina, H., Papakonstantinou, Y., Quass, D., Rajaraman, A., Sagiv, Y., Ullman, J., Vassalos, V., and Widom, J. (1995). "The TSIMMIS Approach to Mediation: Data Models and Languages," *Next Generation Information Technologies and Systems*, Naharia, Israel, November, Anonymous ftp to db.stanford.edu/pub/papers/mis-models-languages.ps.

[Guha91] Guha, R.V. (1991). "Contexts: a formalization and some applications," *Technical Report STAN-CS-91-1399-Thesis*, Department of Computer Science, Stanford University.

[Jako96] Jakobisiak, M. (1996). "Programming the web -- design and implementation of a multidatabase browser," *Technical Report CISL WP #96-04*, Sloan School of Management, Massachusetts Institute of Technology.

[JS95] Jonker, W. and Schuetz, H. (1995). "The ECRC multidatabase system," In *Proc ACM SIGMOD*, page 490.

[KS91] Kim, W. and Seo, J. (1991). "Classifying Schematic and Data Heterogeneity in Multidatabase Systems," *IEEE COMPUTER*, December: 12-18.

[KuL88] Kuhn, E. and Ludwig, T. (1988). "VIP-MDBMS: A logic multidatabase system," In *Proc Int'l Symp. on Databases in Parallel and Distributed Systems*.

[LaR82] Landers, T. and Rosenberg, R. (1982). "An overview of Multibase," In *Proc 2nd International Symposium for Distributed Databases*: 153-183.

[LeG89] Lenat, D.B. and Guha, R. (1989). *Building large knowledge-based systems: representation and inference in the Cyc project*. Addison-Wesley Publishing Co., Inc.

[LFG*97] Lu, H., Fan, W., Goh, C.H., Madnick, S., Cheung, D.W. (1997) "Discovering and Reconciling Semantic Conflicts: A Data Mining Perspective", *Proceedings of the IFIP 2.6 Conference on Database Semantics*, October.

[LFG*98] Lu, H., Fan, W., Goh, C.H., Madnick, S. and Cheung, D.W. (1998) "Discovering and Reconciling Semantic Conflicts: A Data Mining Perspective", *Data Mining and Reverse Engineering: Searching for Semantics,* Chapman & Hall.

[Lit92] Litwin, W. (1992). "O*SQL: A language for object oriented multidatabase interoperability," In Hsiao, D.K., Neuhold, E.J., and Sacks-Davis, R., editors, *Proc of the IFIP WG2.6 Database Semantics Conference on Interoperable Database Systems (DS-5),* Lorne, Victoria, Australis. North-Holland: 119-138.

[LitA87] Litwin, W. and Abdellatif, A. (1987). "An overview of the multi-database manipulation language MDSL," *Proc of the IEEE*, **75**(5): 621--632.

[LM96] Lee, J. and Madnick, S. (1996) "Resolving Context Heterogeneity: Applying Insights from Semantics and Ontology", *Proceedings of the Workshop on Information Technology & Systems*, December.

[LMS96a] Lee, J., Madnick, S. and Siegel, M. (1996) "The Context Interchange Approach to Large Scale Data Integration: Benefits and Challenges", *Proceedings of the International Conference on Data Engineering*, August.

[Mac88] MacGregor, R. (1988). "A Deductive Pattern Matcher," In Proceedings of AAAI-88, The National Conference on Artificial Intelligence. St. Paul, MN, August.

[Mad00] Madnick, S. (2000) "Chapter 6: The MIT Context Interchange Project" in *Data Quality*, Ed: R.Y. Wang, M. Ziad, and Y.W. Lee, Kluwer Academic Publishers: 79-92.

[Mad97a] Madnick, S. (1997) "Database in the Internet Age", *Database Programming & Design*, **10**(1), January: 28-33.

[Madn96] Madnick, S.E. (1996). "Are we moving toward an information superhighway or a Tower of Babel? The challenge of large-scale semantic heterogeneity," In *Proc of the IEEE International Conference on Data Engineering*, pages 2-8. Also reprinted in 21st (a Web 'Zine at http://www.vxm.com/), Vol 1, No. 6, April 1996, pp. Speed 1-10.

[MMS01a] Moulton, A., Madnick, S., and Siegel, M. (2001) "Cross-Organizational Data Quality and Semantic Integrity: Learning and Reasoning about Data Semantics with Context Interchange Mediation", *Proceedings of the Americans Conference on Information Systems* (AMCIS, Boston), August.

[MMS01b] Moulton, A., Madnick, S., Siegel, M. (2001) "Knowledge Representation Architecture for Context Interchange Mediation: Fixed Income Securities Investment Examples", *Proceedings of the First International Workshop on Electronic Business Hubs:* XML, Metadata, Ontologies, and Business Knowledge on the Web (WEBH; Munich, Germany), published by IEEE, September: 50-54.

[MMS98] Moulton, A., Madnick, S. and Siegel, M. (1998) "Context Interchange on Wall Street", *Proceedings of the International Conference on Cooperative Information Systems* (CoopIS'98), September: 271-279.

[MWF*01] Madnick, S., Wang, R., Dravis, F. and Chen, X. (2001) "Improving the Quality of Corporate Household Data: Current Practices and Research Directions", *Proceedings of the Sixth International Conference on Information Quality* (IQ2001, Cambridge), November: 92-104.

[PFPS92] Patil, R.S., Fikes, R.E., Patel-Schneider, P.F., Mckay, D., Finin, T., Gruber, T., and Neches, R. (1992). "The DARPA Knowledge Sharing Effort: progress report," In *Principles of Knowledge Representation and Reasoning: Proc of the Third International Conference*, Cambridge, MA.

[PGM95] Papakonstantinou, Y. and Garcia-Molina, H. (1995). "Object Fusion in Mediator Systems (Extended Version)," Anonymous ftp to db.stanford.edu/pub/papakonstantinou/1995/fusion-extended.ps.

[PGMU96] Papakonstantinou, Y., Garcia-Molina, H., and Ullman J. (1996). "MedMaker: A Mediation System Based on Declarative Specifications," *International Conference on Data Engineering*, New Orleans, LA, February p. 132-41. Anonymous ftp to db.stanford.edu/pub/papers/medmaker. ps.

[PGMW95] Papakonstantinou, Y., Garcia-Molina, H., and Widom, J. (1995). "Object exchange across heterogeneous information sources," In *Proc IEEE International Conference on Data Engineering*.

[QRS95] Quass, D., Rajaraman, A., Sagiv, Y., Ullman, J., and Widom, J. (1995). "Querying semistructured heterogeneous information," In *Proc International Conference on Deductive and Object-Oriented Databases*.

[Shum96] Shum, A. (1996). *Open Database Connectivity of the Context Interchange System*, Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.

[SL90] Sheth, A.P. and Larson, J.A. (1990). "Federated database systems for managing distributed, heterogeneous, and autonomous databases," *ACM Computing Surveys*, **22**(3): 183--236.

[SM91b] Siegel, M. and Madnick, S. (1991) "A metadata approach to solving semantic conflicts," In *Proc of the 17th International Conference on Very Large Data Bases*: 133-145.

[SSR94] Sciore, E., Siegel, M., and Rosenthal, A. (1994). "Using semantic values to facilitate interoperability among heterogeneous information systems," *ACM Transactions on Database Systems*, **19**(2): 254-290.

[SYE90] Scheuermann, P., Yu, C., Elmagarmid, A., Garcia-Molina, H., Manola, F., McLeod, D., Rosenthal, A., and Templeton, M. (1990). "Report on the workshop on heterogeneous database systems," *ACM SIGMOD RECORD*, 19(4): 23--31. Held at Northwestern University, Evanston, Illinois, Dec 11-13, 1989. Sponsored by NSF.

[TBD87] Templeton, M., Brill, D., Dao, S.K., Lund, E., Ward, P., Chen, A. L.P., and MacGregor, R. (1987). "Mermaid --- a front end to distributed heterogeneous databases," *Proc of the IEEE*, **75**(5): 695-708

[TM97] Tu, S.Y., Madnick S. (1997) "Incorporating Generalized Quantifiers into Description Logic for Representing Data Source Contents", *Proceedings of the IFIP 2.6 Conference on Database Semantics*, October.

[TRV95] Tomasic, A., Raschid, L., and Valduriez, P. (1995). "Scaling heterogeneous databases and the design of DISCO," In *Proc of the 16th International Conference on Distributed Computing Systems*, Hong Kong.

[ZMS01] H. Zhu, S. Madnick, and M. Siegel, "Information Aggregation - a Value-Added E-Service", Proceedings of the 5th International Conference on Technology, Policy, and Innovation – Theme: "Critical Infrastructures", Delft, The Netherlands, June 26-29, 2001.

International Relations (IR)

[AK97] Arbetman, M., and Kugler, J. eds. (1997). Political Capacity and Economic Behavior. Boulder, CO, Westview Press.

[AM99] Amadae, S. M., De Mesquita, B. B. (1999). The Rochester School: The Origins of Positive Political Theory. Annual Review of Political Science. Nelson W. Polsby, ed. Palo Alto, CA, Annual Reviews. **2:** 269-295.

[Bate01] Bates, R. H. (2001). Prosperity and Violence - The Political Economy of Development. New York, London, W. W. Norton and Company.

[BHP92] Bright, M., Hurson, A., and Pakzad, S. (1992). "A taxonomy and current issues in multidatabase systems," *IEEE Computer*, 25(3):50--60.

[BJ99] Becker, E., and Jahn, T. eds. (1999). Sustainability and the Social Sciences. London, Zed Books, Ltd.

[Brech96] Brecher, M. (1996). "Introduction: Crisis, Conflict, War - State of the Discipline." International Political Science Review **17**(2): 127-139.

[Cho81] Choucri, N. (1981). International Energy Futures - Petroleum Prices, Power and Payments. Cambridge MA, The MIT Press.

[Cho92] Choucri, N. (1992). "Environment and Conflict: New Principles for Environmental Conduct." Disarmamemnt **XV**(1): 67-78.

[Cho93a] Choucri, N. (1993). "Political Economy of the Global Environment." International Political Science Review **14** (1): 103-116.

[Cho93b] Choucri, N. (1993). Environmentalism. Oxford Companion to Politics of the World. e. J. Krieger. New York, Oxford University Press.

[Cho93c] Choucri, N., ed. (1993). Chapter 1: Introduction: Theoretical, Empirical, and Policy Perspectives. Global Accord: Environmental Challenges and International Responses. Cambridge, Mass., MIT Press.

[Cho95] Choucri, N. (1995). "Globalization of Eco-Efficiency: GSSD on the WWW." UNEP Industry and Environment(October-December 1995): 45-49.

[Cho98] Choucri, N., McHugh and Millman (1998). Innovations in Cyberpartnering for Sustainability. Care Innovation '98: Proceedings, Second International Symposium, Brokerage Event and Environmental Exhibition, Vienna, Austria.

[Cho99a] Choucri, N. (1999). Innovations in Use of Cyberspace. Sustainability and the Social Sciences. Becker and Jahn, eds. London, Zed Books, Ltd.**:** 274-283.

[Cho99b] Choucri, N. (1999). Strategic Partnerships with Multilingual Functionality for Globalisation and Localisation. European Commission Directorate - General Information Society Workshop on "Sustainability and Environment", IST99, Helsinki, Finland.

[Cho99c] Choucri, N. (1999). The Political Logic of Sustainability. Sustainability and the Social Sciences. Becker and Jahn, eds. New York, Zed Books.

[ChoMM98] Choucri, N., McHugh and Millman (1998). Innovations in Cyberpartnering for Sustainability. Care Innovation '98: Proceedings, Second International Symposium, Brokerage Event and Environmental Exhibition, Vienna, Austria.

[ChoN75] Choucri, N., North, R. C. (1975). Nations in Conflict - National Growth and International Violence. San Francisco, CA, W. H. Freeman and Company.

[ChoN92] Choucri, N., North, R. C., and Yamakage S. (1992). The Challenge of Japan - Before World War II & After. New York, Routledge. [CHS91a] Collett, C., Huhns, M., and Shen, W. (1991). "Resource Integration Using an Existing Large Knowledge Base," *MCC Technical Report Number ACT-OODS-127-91*.

[Elki97] Elkins, D. J. (1997). "Globalization, Telecommunication, and Virtual Ethnic Communities." International Political Science Association **18**(2): 139-152.

[Fark96] Farkas, A. (1996). Evolutionary Models in Foreign Policy Analysis. International Studies Quarterly Hart, Rasler, Thompson, eds. Cambridge, MA, Blackwell Publishers, Inc. **40:** 343-362.

[Gilp96] Gilpin, R. (1996). Economic Evolution of National Systems. International Studies Quarterly. Hart, Rasler, Thompson, eds. Cambridge, MA, Blackwell Publishers, Inc. **40:** 411-432.

[HSG80] Holsti, O.R., Siverson, R.M., and George, A.L. eds. (1980). Change in the International System. Boulder, CO. Westview Press.

[HT99] Herst, P and Thompson, G. (1999). Globalization in Question. Malden, MA, Blackwell Publishers Inc.

[Jerv78] Jervis, R. (1978). "Cooperation under the Security Dilemma." World Politics: 167-214.

[KR96] Kang, H, and Reuveny R. (1996). International Conflict and Cooperation: Splicing COPDAB and WEIS Series. International Studies Quarterly. Hart, Rasler, Thompson, eds. Cambridge, MA, Blackwell Publishers, Inc. **40:** 281-306.

[Mesq00] Mesquita, B. B. (2000). Principles of International Politics - People's Power, Preferences, and Perceptions. Washington, D.C., CQ Press.

[Midl89] Midlarsky, M. I. ed. (1989). Handbook of War Studies, The University of Michigan Press.

[Mode96] Modelski, G. (1996). Evolutionary Paradigm for Global Politics. International Studies Quarterly. Hart, Rasler, Thompson, eds. Cambridge, MA, Blackwell Publishers, Inc. **40:** 321-342.

[NC93] North, R. C. and Choucri, N. (1993). Growth, Development, and Environmental Sustainability : Profiles and Paradox. Global Accord: Environmental Challenges and International Responses. N. Choucri., ed. Cambridge, MA, MIT Press.

[ND00] Nye, J. S. and Donahue, J. eds. (2000). Governance in a Globalizing World. Washington, D.C., Brookings Institution Press.

[OY96] Osherenko, G. and Young, O. (1996). "Testing Theories of Regime Formation - Findings from a Large Collaborative Research Project." The International Political Economy and International Institutions **1**: 573-602.

[PM96] Poznanski, K. and Modelski, G. (1996). Evolutionary Paradigms in the Social Science. International

Studies Quarterly. Hart, Rasler, Thompson, eds. Cambridge, MA, Blackwell Publishers, Inc. **40:** 315-319.

[SK92] Sheth, A. and Kashypa, V. (1992). "So Far (Schematically) yet So Near (Semantically)," in Hsiao, D., Neuhold, E., and Sacks-Davis R., ed. *Proc of the IFIP WG2.6 Database Semantics Conference on Interoperable Database Systems (DS-5)*, p. 283-312.

[Stei00] Steinbruner, J. D. (2000). Principles of Global Security. Washington, D.C., Brookings Intitution Press.

[Vasq93] Vasquez, J. A. (1993). The War Puzzle. Cambridge, Cambridge University Press.

[Walt83] Waltz, K. N. (1983). Theory of International Politics. Reading, MA, Addsion-Wesley Publishing Company.

[Zieg84] Ziegler, D. W. (1984). War, Peace, and International Politics. Boston, Little, Brown and Company.