Information Infrastructure for Archiving & Integrating Primary Archaeological Data

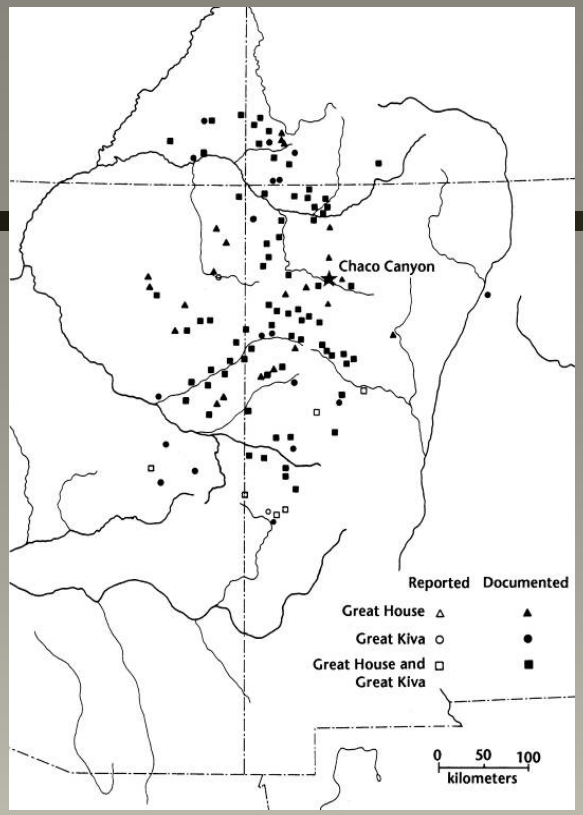
Keith W. Kintigh **kintigh@asu.edu**
School of Human Evolution & Social Change
Global Institute of Sustainability
Arizona State University
Tempe, Arizona 85287-2402, US

Principal Collaborators:
K. Selçuk Candan, Hasan Davulcu, Subbarao Kambhampati
Jane Buikstra, Ben Nelson, Margaret Nelson, Katherine Spielmann

ASU Arizona State University

School of Human Evolution & Social Change
Department of Computer Science & Engineering
Global Institute of Sustainability

Chacoan Outliers



Crown & Judge

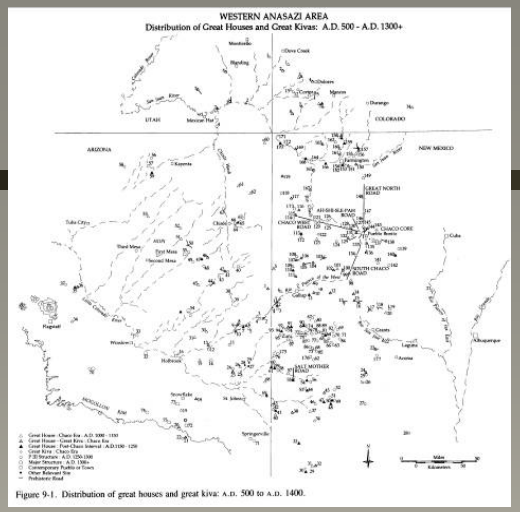


Figure 9-1. Distribution of great houses and great kiva: A.D. 500 to A.D. 1400.

Fowler & Stein

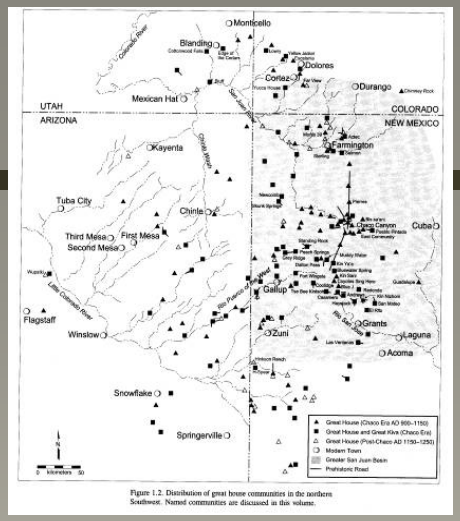
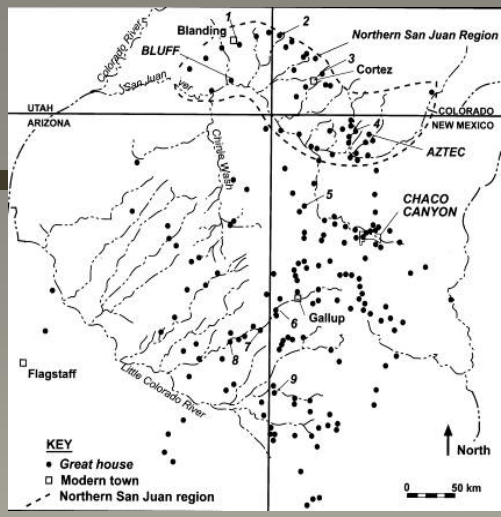


Figure 1.2. Distribution of great house communities in the northern Southwest. Named communities are discussed in this volume.

Mahoney



Cameron

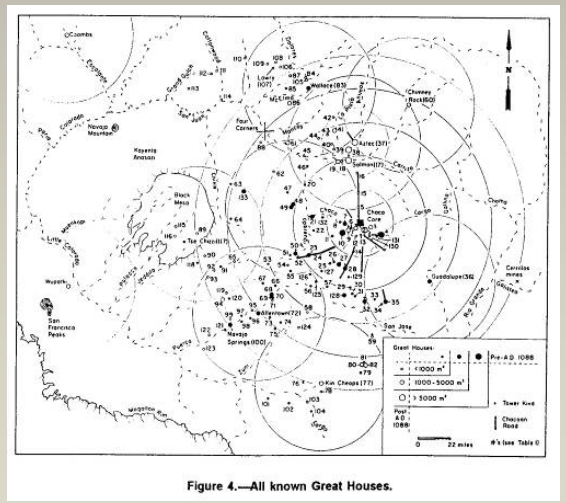


Figure 4.—All known Great Houses.

Wilcox

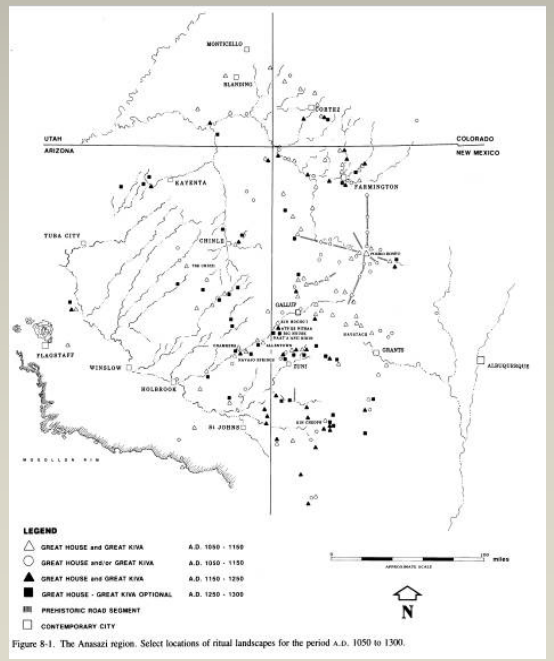
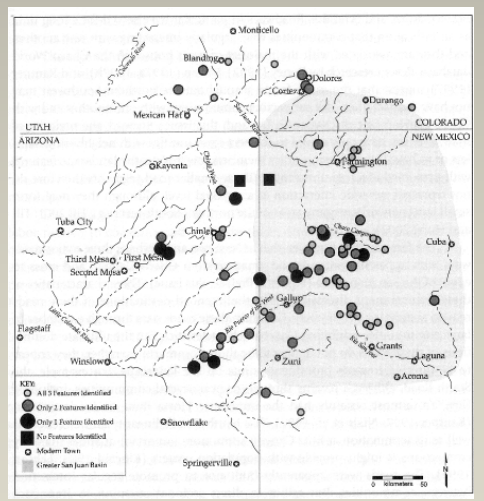


Figure 8-1. The Anasazi region. Select locations of ritual landscapes for the period A.D. 1050 to 1300.

Stein & Lekson



Kantner

ASU Arizona State University

School of Human Evolution & Social Change
Department of Computer Science & Engineering
Global Institute of Sustainability

Why are these Maps Different?

- Problems
  - Inconsistent definitions of Chacoan Outlier
    - What features are necessary and sufficient
    - Limited surface visibility of some features
  - Conflation of Chacoan & post-Chacoan Times
  - Can't sort out differences based on published conclusions
- Solution requires going back to ***primary data***
  - Starting in 1999, the ASU team's interest in synthetic problems has led us to try to advance a "cyberinfrastructure" for archaeology.

Challenges for Synthesis

- Primary data are being irretrievably lost
  - Discard (paper & digital)
  - Media degradation (digital)
  - Software & media obsolescence (digital)
  - Lack of adequate metadata (paper & digital)
- Data integration (constructing analytically comparable observations) is difficult because:
  - Lack of data comparability across projects
  - Limited access to primary data
  - Lack of systematic metadata describing data semantics
  - Structural complexity of archaeological datasets

ASU Arizona State University

School of Human Evolution & Social Change
Department of Computer Science & Engineering
Global Institute of Sustainability

Working toward a Shared Vision

- Developed at a 2004 NSF HSD-Funded Workshop in Santa Barbara w/ 31 Scholars
  - archaeology, physical anthropology, ecology, & computer science
- Vision: An information infrastructure for systematically collected archaeological data to:
  - Sustain the utility of new and legacy data.
  - Provide world-wide access to primary data
  - Enable data integration so that we can do synthetic research *at scales not currently feasible*
  - Foster use of legacy data
- Workshop report
  - *American Antiquity* (July 2006) http://cadi.asu.edu
  - Endorsed by SAA, AAPA, SHA

ASU Arizona State University

School of Human Evolution & Social Change
Department of Computer Science & Engineering
Global Institute of Sustainability

Infrastructure Components

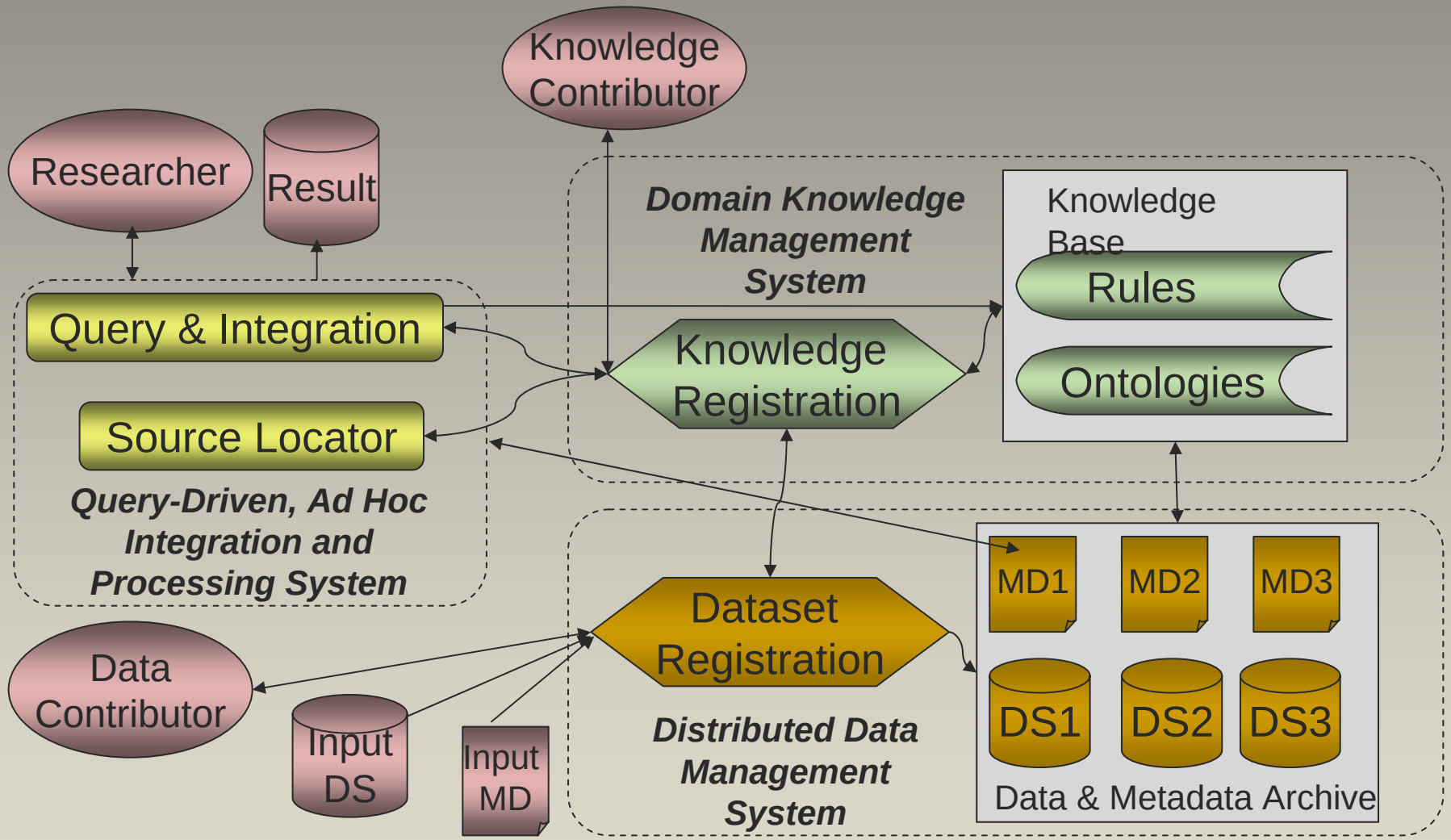
- Software Registration Tools
  - that facilitate the registration of datasets in a central catalog, and
  - foster the collection of adequate metadata.
- Internationally distributed network of data sources
- Web-based, concept-oriented query interface
- Integration system that uses systematized knowledge, digital metadata, and user guidance to integrate data collected
  - at different scales,
  - at different times,
  - by different investigators
  - using variable data recovery strategies, and
  - inconsistent typologies.
- Output databases of scaled and integrated (comparable) observations
  - Useable for further analysis

Technical Approach

- Achieving the vision entails
  - Preserving the original semantics of data
  - Dealing adequately with the complexities of archaeological contexts
  - Resolving conflicts that derive from incomplete and inconsistent data recording strategies.
  - Interactively reconciling the data demands of the query with the semantic content of the datasets

Conceptual Design – KADIS

Knowledge-based Archaeological Data Integration System



ASU Arizona State University

School of Human Evolution & Social Change
Department of Computer Science & Engineering
Global Institute of Sustainability

National Science Foundation
WHERE DISCOVERIES BEGIN

Current NSF Project

- Use faunal data as a proof-of-concept testbed.
  - Not too hard, not too easy; of interest beyond archaeology
  - Work with International Council for Archaeozoology (ICAZ)
- Work with other science informatics initiatives
- Supply tools that facilitate registration of new datasets
- Establish an international, Web-accessible, distributed network of archaeological (fauna) data sources
- Provide an interface that uses *concept-based* queries
- Develop software that uses archaeological *knowledge* to *integrate* information across diverse databases
- Over the Web, return a database of scaled and integrated observations to users

ASU Arizona State University
School of Human Evolution & Social Change
Department of Computer Science & Engineering
Global Institute of Sustainability

Promise of Data Integration

- Provide for long term preservation of archaeological data *and their semantic content.*
- Expand international access to primary archaeological data
- Advance key archaeology research questions
- Massively improve our ability to model long-term stability and change.
- Make archaeological and environmental data useable for research by scholars in other fields
- Contribute data integration tools useful in other sciences whose data share key characteristics with archaeology

ASU
Arizona State University

School of Human Evolution & Social Change
Department of Computer Science & Engineering
Global Institute of Sustainability

ASU Project Team

School of Human Evolution & Social Change (Anthropology)

Keith Kintigh - Archaeology, Cíbola, Political Organization, Settlement Patterns
Katherine Spielmann - Archaeology, North America, Craft Specialization, Fauna
Margaret Nelson - Mimbres Archaeology, Stone Tools, Abandonment
Ben Nelson - Archaeology, Mesoamerica, Complex Societies, Exchange
Jane Buikstra - Bioarchaeology, South America, Mesoamerica
Michelle Hegmon - Archaeology, Mesa Verde/Mimbres, Social Theory, Style
Marty Anderies – Mathematics, Systems Modeling, Resilience
Sander van der Leeuw – Archaeology, Mediterranean, Modeling, Ceramics

School of Computing and Informatics

K. Selcuk Candan - Heterogeneous Data Management & Integration
Hasan Davulcu – Heterogeneous Data Sources, Data Mining
Subbarao Kambampati –Data Integration, Artificial Intelligence Query Processing
Chitta Baral – Knowledge Representation, Imprecise Data Management
Huan Liu - Ontology Management, Data Mining