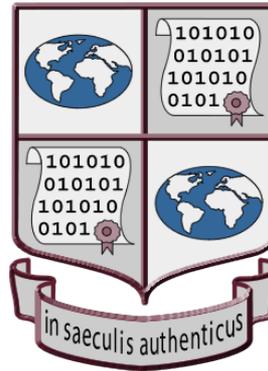


InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems



**CS14 – Archaeological Records in a GIS:
Research in the American Southwest**

**GS09 – Survey of Record-keeping Practices
of GIS Archaeologists**

CLAID Team Workshop, Vancouver, BC
17 November 2005

Randy Preston

School of Library, Archival and Information Studies
The University of British Columbia, Canada



InterPARES 2 Project

Randy Preston

CS14/GS09 – CLAUD Team Workshop, 17 Nov 2005

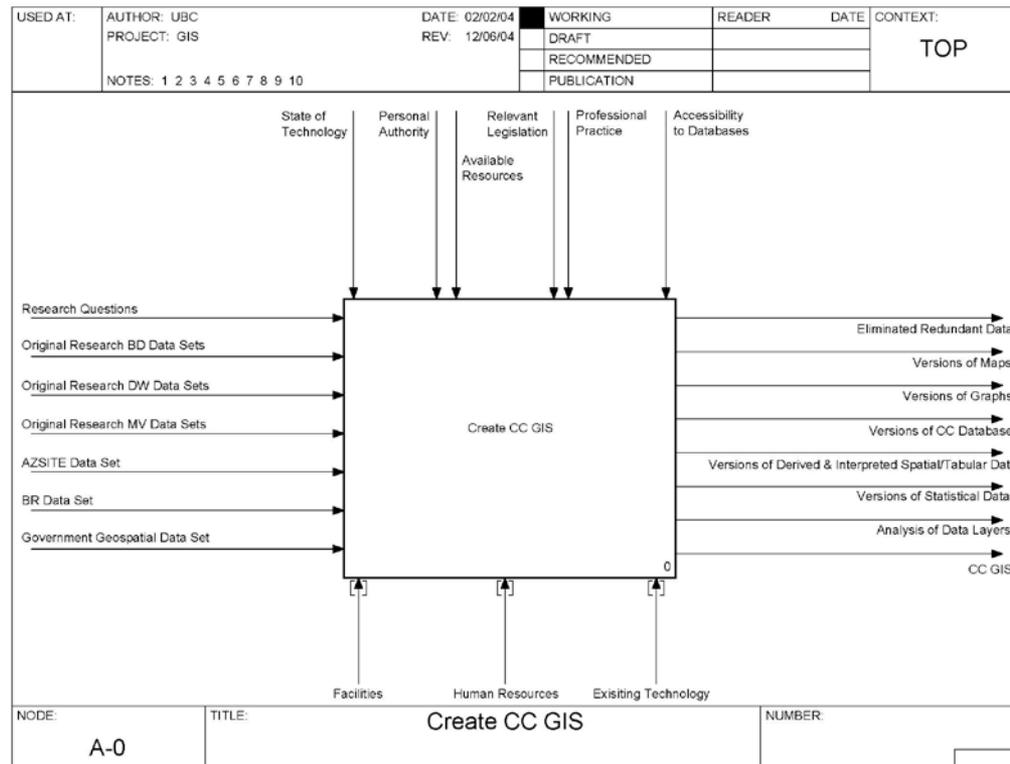
Data Collection Methods

- 1. Case Study Interviews (5)**
 - 2. Online GIS Survey**
 - 3. Review of existing literature**
- 23 questions in case study report focus on the interview data
 - Survey and literature review data used to help assess the representativeness or generalisability of the interview data



Case Study 14

Archaeological Records in a GIS: Research in the American Southwest



Provenancial Context

- Looked at recordkeeping activities associated with Coalescent Communities (CC) Database and GIS created by Center for Desert Archaeology (CDA) in Tucson, AZ <http://www.centerfordesertarchaeology.org/>
- Private, non-profit research organization whose primary mandate is to promote:

“the stewardship of archaeological and historic resources in the American Southwest and Mexican Northwest through active research, preservation, and public education”



Provenancial Context

- CDA employs one full-time archaeologist as a “GIS Specialist” – primary responsibility is development and management of the CC Database/GIS
- CC Database/GIS is one component of the larger *Coalescent Communities in Arizona Project (A.D. 1200-1540)*
- Project goal is to research the aggregation and migration of peoples in the prehistoric Southwest



Juridical-Administrative Context

- Actions of archaeologists in the US are affected by a complex web of legal and regulatory considerations that vary from state to state
- Also affected by various ethical and professional guidelines and standards



Procedural Context

- Few, if any, formal documentary procedures were identified
- No procedural manuals exist for creation, management, use or preservation of the CC GIS
- GIS Specialist is the “procedures manual”
- Establishment of formal documentary procedures seen as excessive and unrealistic in relation to the resources allocated to the project and the nature of the organization



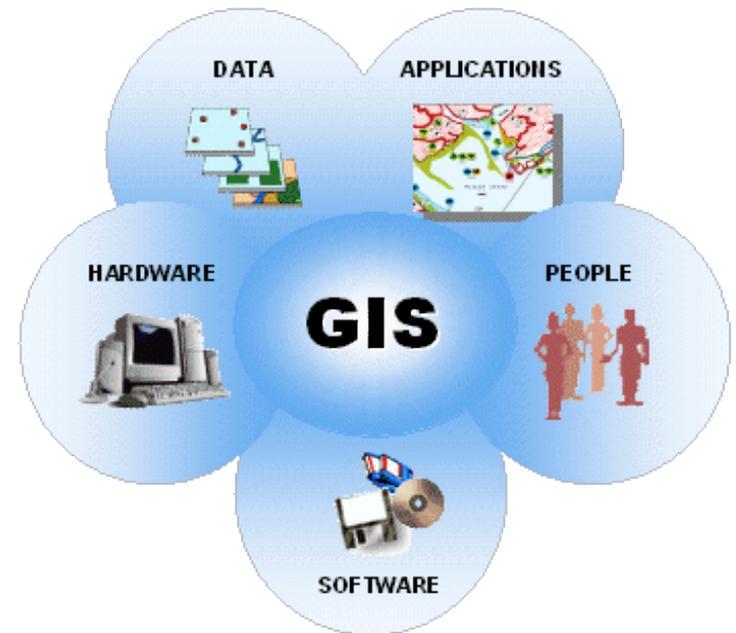
Documentary Context

- CC Project records are a series within the CDA funds
- Majority of CDA funds relates to specific projects and research grants awarded to the organization
- No traditional recordkeeping elements, such as formal classification schemes, record inventories, indexes or registers, were identified associated with the activities of the CDA



Technological Context

- System incorporates common office technologies, including personal computers networked to a central file server (LAN)
- No custom hardware or software identified
- Current and non-current data typically stored on the LAN, but may also reside on the GIS Specialist's personal computer as well as on CD-Rs



Technological Context

- Generalized application software used:
Windows 2000, MS Office Suite (esp. Access, Excel, Word), EndNote, Adobe Acrobat, Topo! (topo. map software), Notepad, and unspecified basic Microsoft image editing/analysis software
- Specialized application software used:
ESRI ArcView, versions 3.2 and 8.2, including statistical analysis add-ins



CC Database

- Microsoft Access database
- Consists of 5 sets of pre-existing archaeological site data collected from researchers, repositories and literature searches
- Data sets from multiple formats, including paper, digital spreadsheets and other pre-existing computer databases



General Study 09

Survey of Record-keeping Practices of GIS Archaeologists

InterPARES Project: - Opera

File Edit View Navigation Bookmarks Mail Chat Tools Window Help

Rewind Back Forward Reload New Open Save Print preview Fullscreen Print Tile Security

New page InterPARES Project:

http://www.interpares.org/cs14/cs14_survey.cfm?q=02&CFID=29947&CFTOKEN=65384069 Go Google search

InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

survey of record-keeping practices of archaeologists

B. GIS EXPERIENCE/BACKGROUND

B1. How long have you been using GIS?

- Less than 1 year
- 1 to 2 years
- 2 to 3 years
- 3 to 4 years
- 4 to 5 years
- 5 to 6 years
- 6 to 7 years
- 7 to 8 years
- 8 to 9 years
- 9 to 10 years
- More than 10 years
- Not sure

Comments - Use this space if you wish to comment further on this question:

● Please answer carefully, answers can only be submitted once. ●

Next question

Question 02 of 39

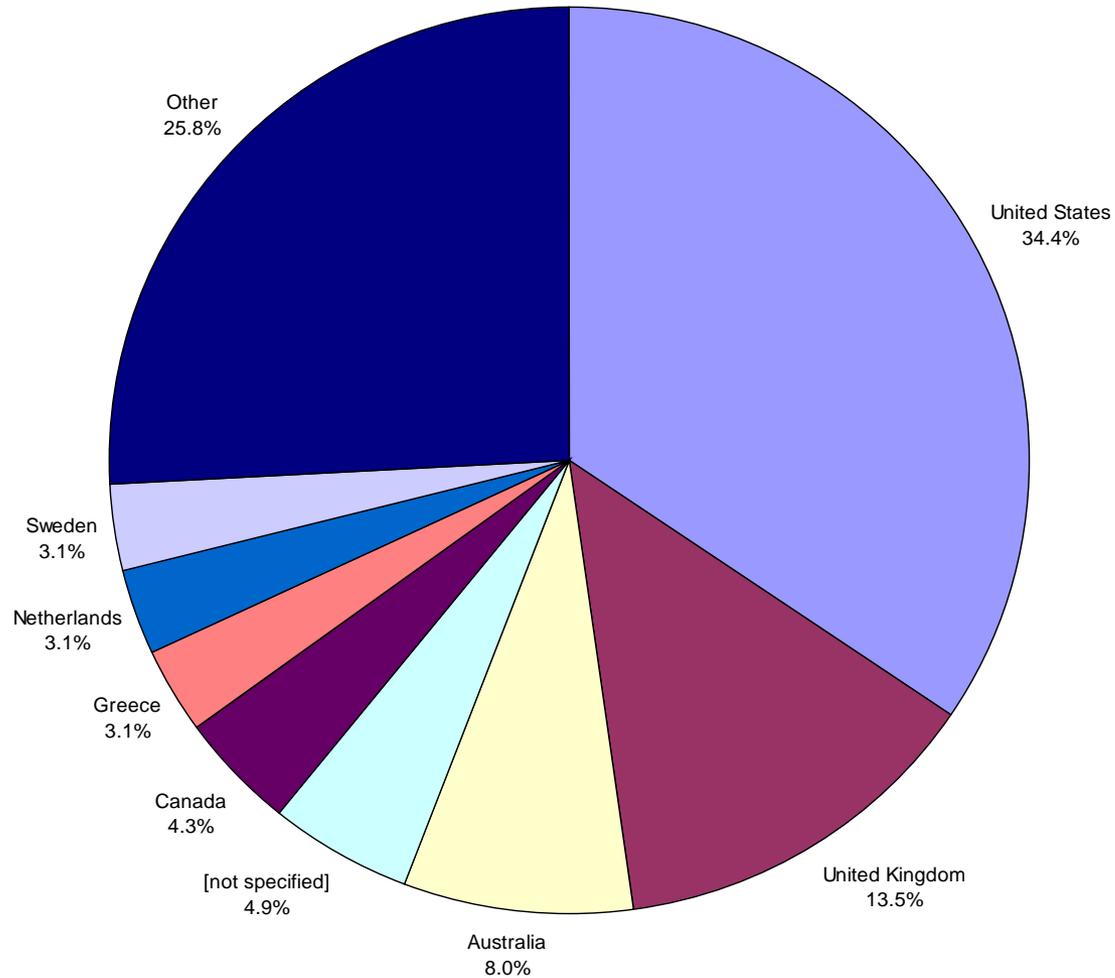


GIS Survey Questionnaire

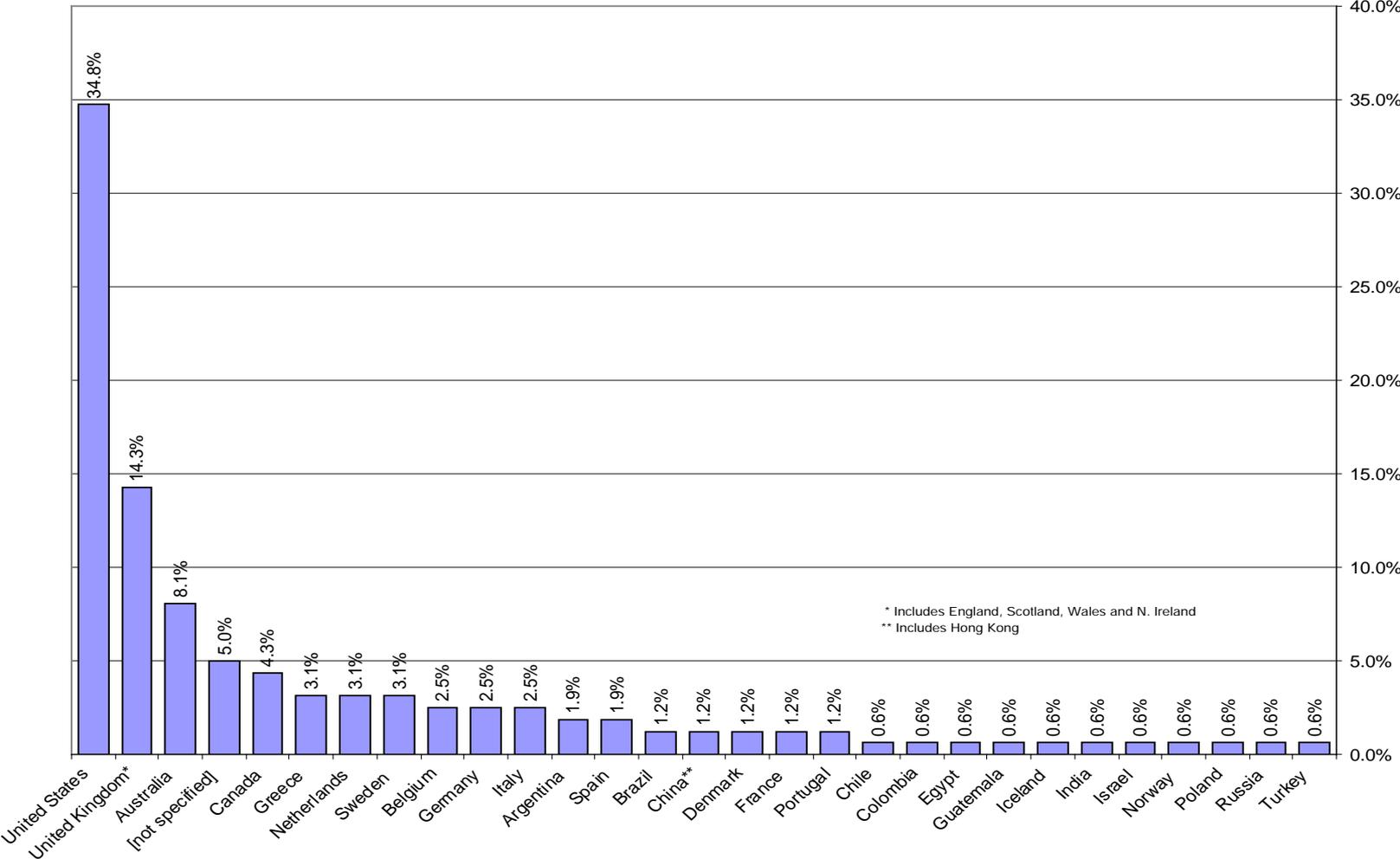
- 40 questions that addressed specific records creation, documentation, management and preservation issues identified during case study interviews
- 900+ GIS archaeologists from 69 countries worldwide invited to participate
- Online for one month
http://www.interpares.org/cs14/login_rp.cfm
- 157 usable responses received from archeologists in 30 countries



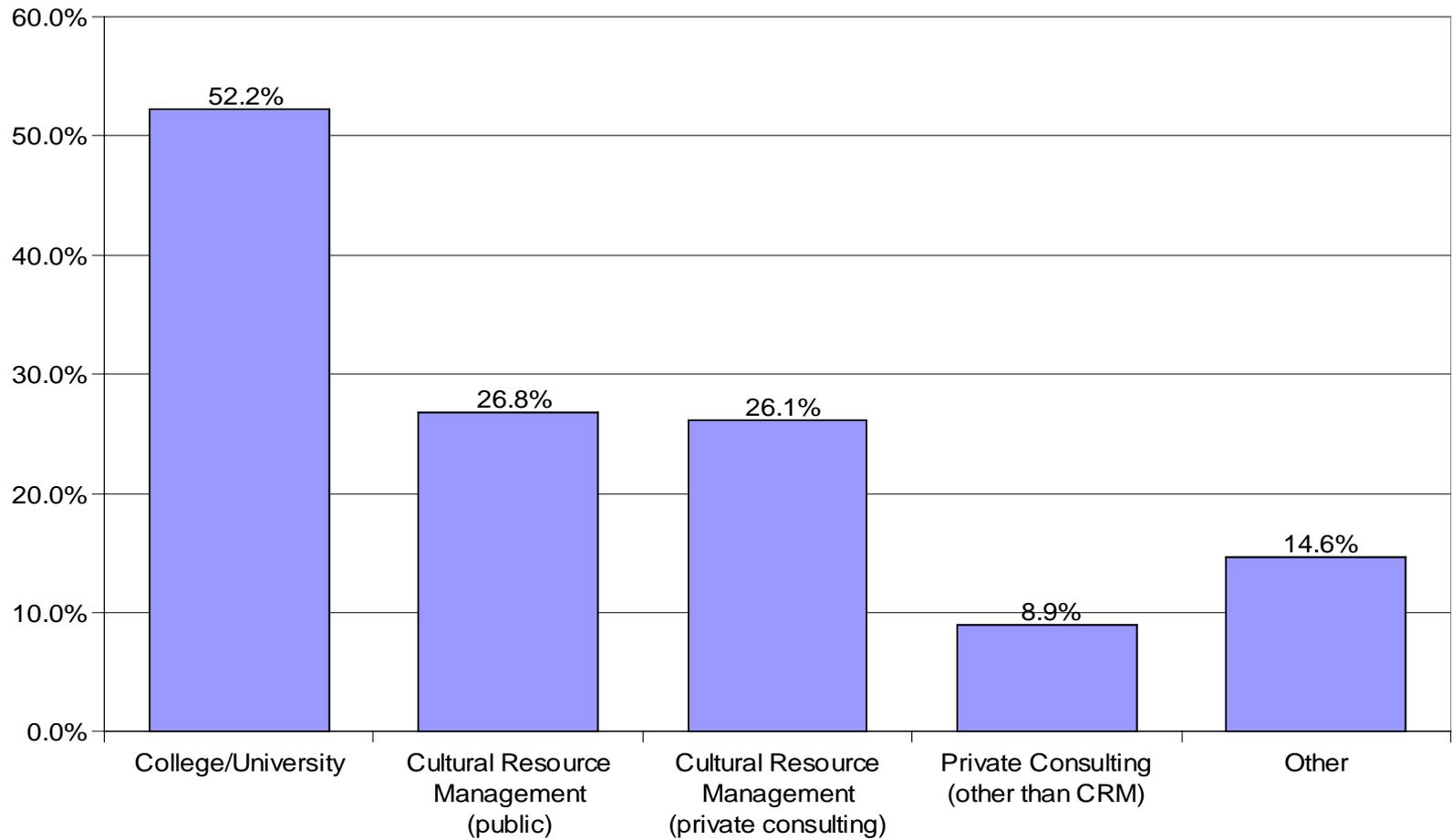
Survey Participants by Country



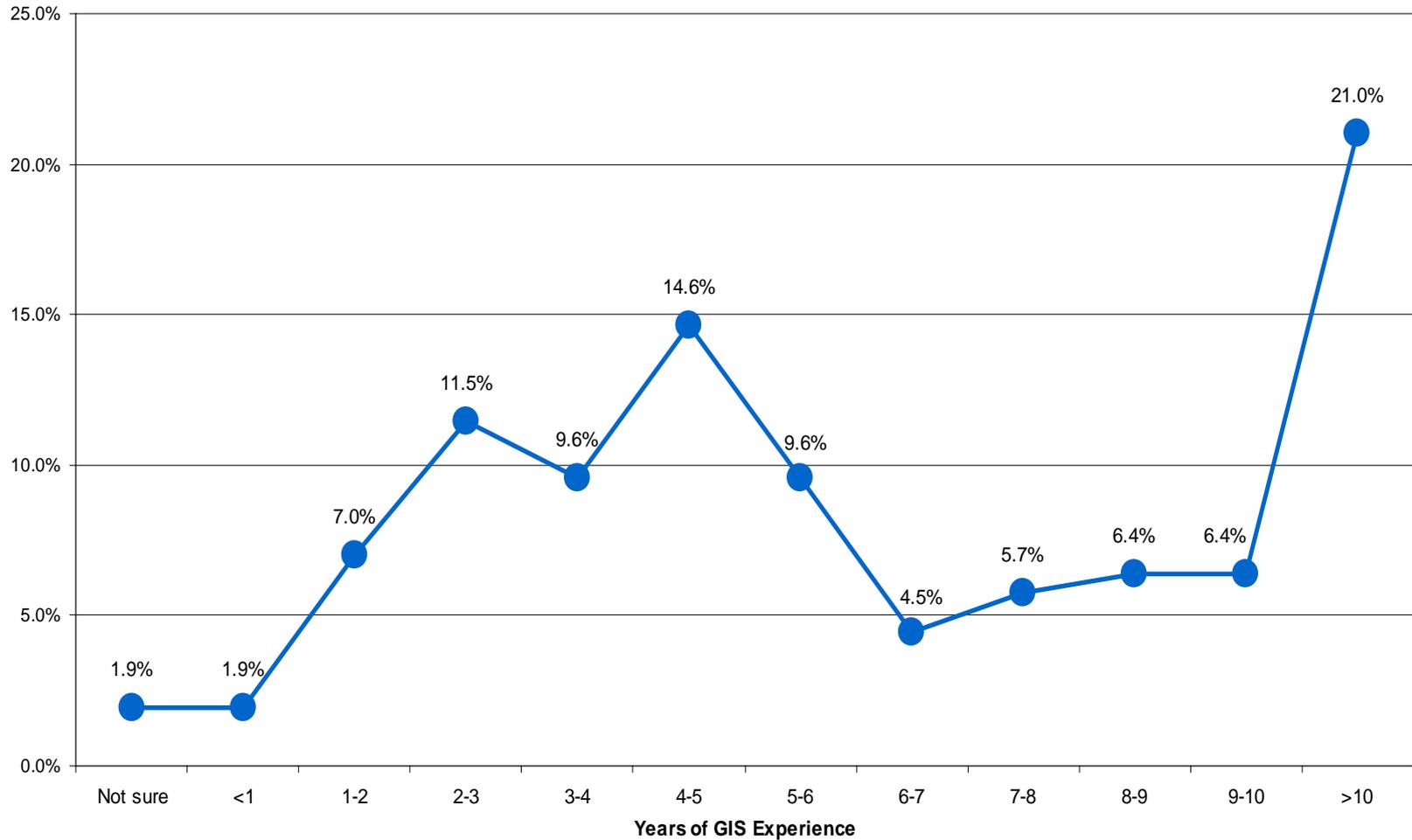
Survey Participants by Country



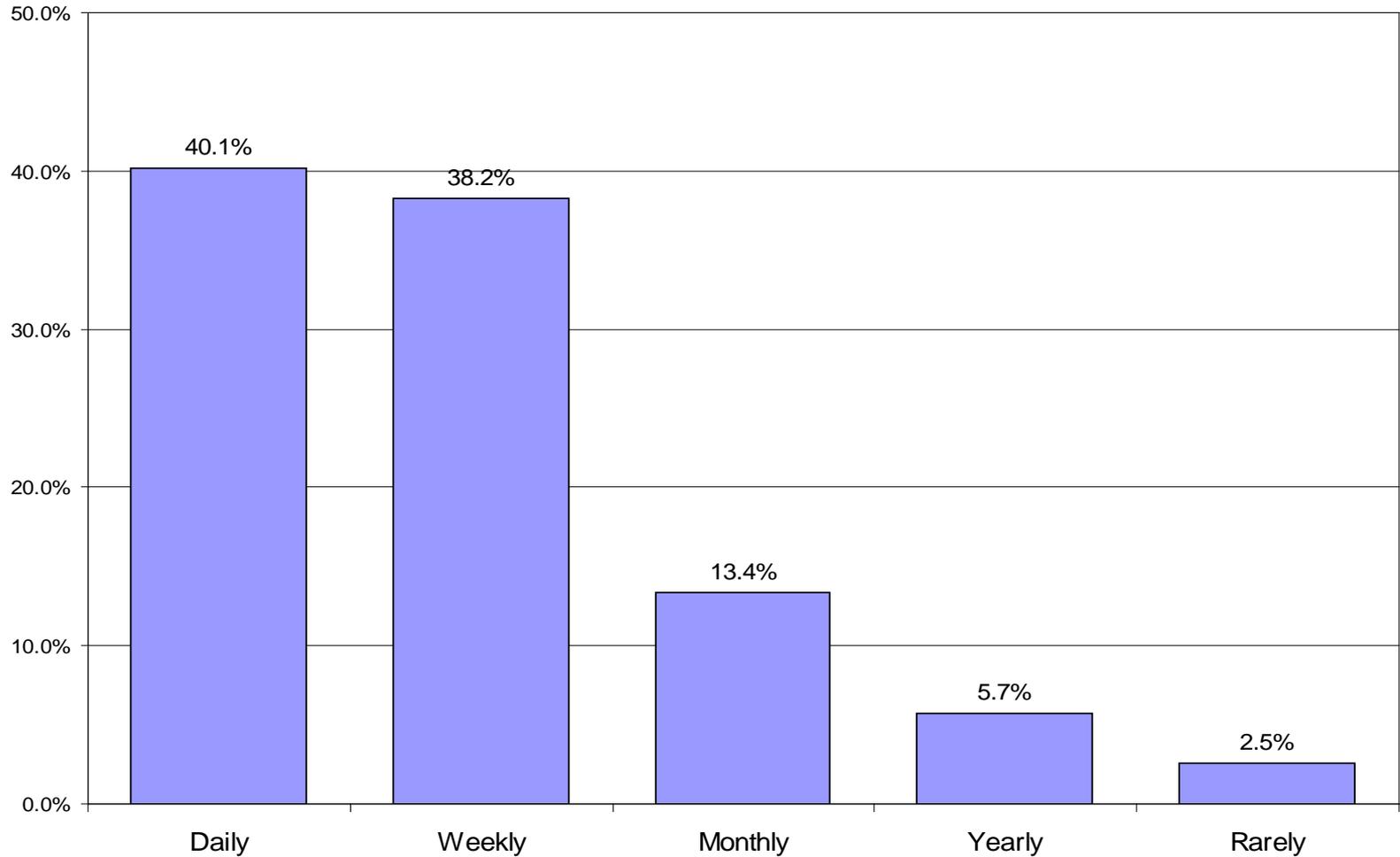
Professional Affiliation



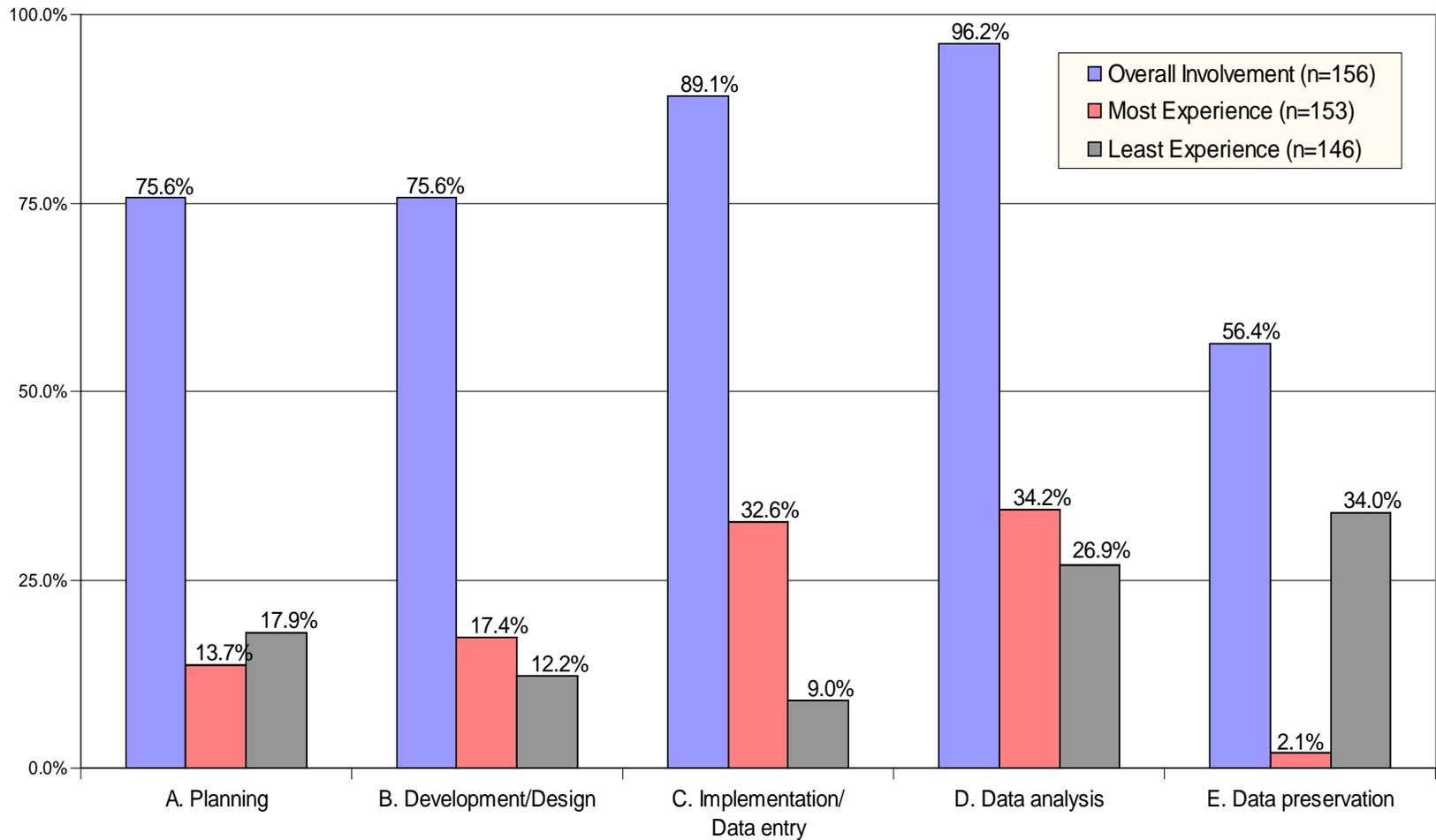
Years of GIS Experience



Frequency of GIS Use



GIS Component Experience



Survey Sections

- A. Introduction (1)
- B. GIS Experience/Background (7)
- C. File Management/Documentation Practices (11)
- D. Digital Preservation Practices (13)
- E. Data Input/Output Practices (2)
- F. Record Quality, Reliability & Authenticity Issues (5)
- G. General Comments (1)



A. Introduction

- 1 Question
- Free-text question asking participants to provide a brief definition of a GIS, including what they considered to be its most important, significant, and/or distinguishing components and functions.
- Intended primarily to assess what importance, if any, participants assigned to their role (i.e., the human operator) in the overall geographic information system.



B. GIS Experience/Background

- 7 Questions
- Intended primarily to gather basic background information about the participants and their current level of experience with GIS projects, including how long they have been using GIS, how often they use it, where they use it, etc.
- This information was used for statistical purposes (e.g., correlate experience with record-keeping procedures used)



C. File Mgmt/Documentation Practices

- 11 Questions
- Intended primarily to assess the participants' GIS project and file documentation and management habits.
- Participants were asked, for example, about their file naming procedures, the ways in which they managed file version control and how often they created non-electronic records associated with their electronic records.
- Of particular interest here was identifying the ways in which the participants dealt with modifications to their data files in terms of if, when, and how they typically documented these changes.



D. Digital Preservation Practices

- 13 Questions
- Intended primarily to gather information about the preservation strategies used by the participants when saving their GIS projects for the long-term, either “in-house” or in a designated repository.
- In addition, a number of questions dealt with the issue of metadata, looking especially at how the participants recorded metadata, what, if any, standards they followed, and how they integrated metadata into their GIS projects.



E. Data Input/Output Practices

- 2 Questions
- Intended primarily to assess the overall routineness of the procedures used when creating and manipulating the various components and outputs of GIS projects, and the degree to which those procedures are based on any sort of GIS procedures manual.



F. Record Quality, Reliability and Authenticity Issues

- 5 Questions
- Focused primarily on data/record reliability, authenticity and accuracy issues related to the participants' GIS projects.
- Asked to identify what, if any...
 - significance the concept of “accuracy” held for them
 - record/data auditing practices are used in their projects
 - procedures are used to identify authors/creators
 - security measures are used to control access to, and prevent unauthorized modifications to, project files.

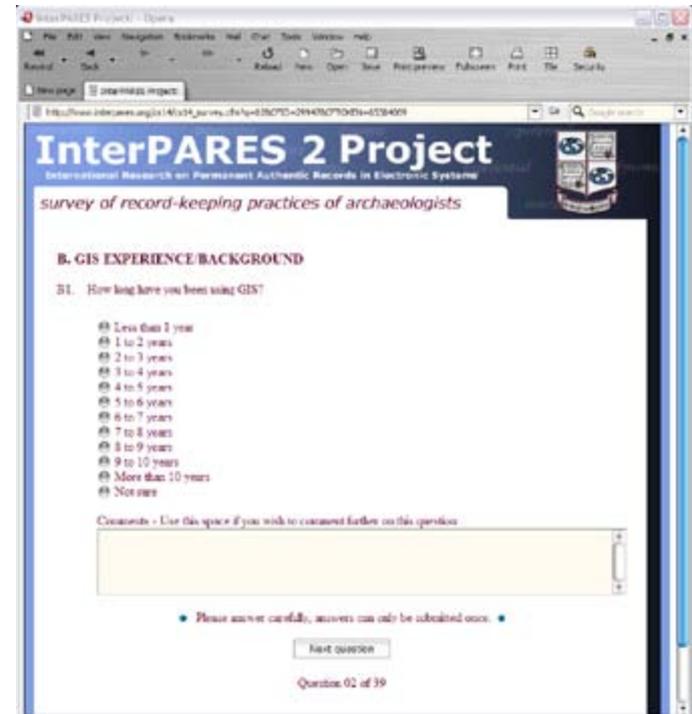
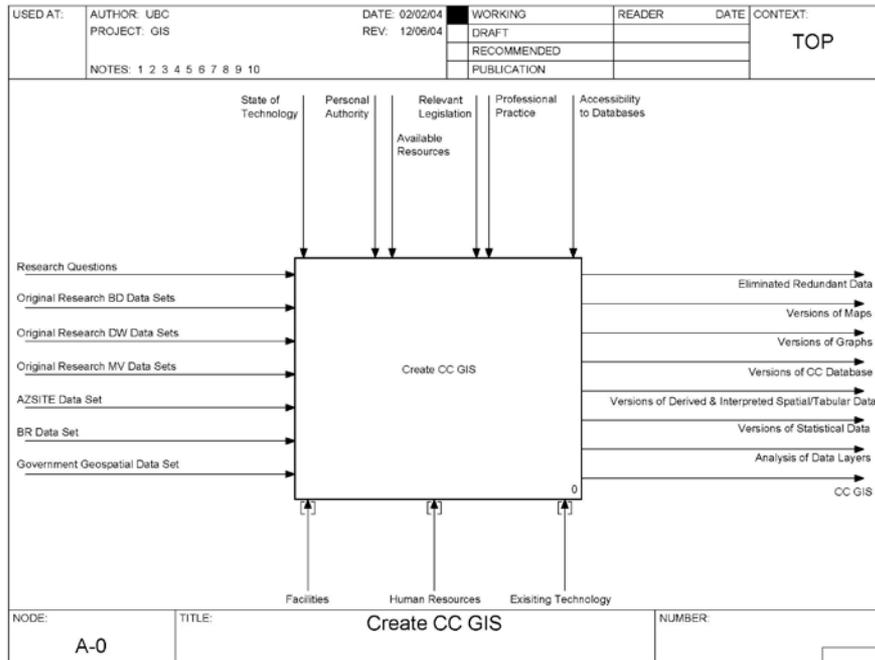


G. General Comments

- 1 Question
- The final section was optional and consisted of a free-text question asking participants to submit any final thoughts about their GIS record-keeping activities or experiences that they thought might be pertinent to the survey.



Case Study & Survey Results



Records Creation, Documentation and Management Issues

Case Study Results

- No part of CC GIS intentionally designed to function as record-making or record-keeping system in the sense assumed by the MCP model
- GIS Creation, maintenance, file naming and versioning procedures all largely ad hoc, idiosyncratic and undocumented
- Any documentation that occurs almost invariably is to facilitate replication of likely-to-be-repeated analysis procedures, rather than to serve any records management function



Records Creation, Documentation and Management Issues

Case Study Results

- No formal, systematic or documented records classification scheme
- No deliberate, creator-imposed metadata schema, only “incidental” metadata
- Considering creating proprietary metadata schema to document data source information (author, date of recording, etc.) for use with ArcCatalogue, a metadata tool in new version of ArcView



Records Creation, Documentation and Management Issues

Survey Results

- 83% of survey participants routinely create documentation for their records
- however, documentation procedures tend to be informal for 59% of these participants
- documentation usually occurs at or near the end of projects for 25% of participants
- only 37% create documentation concurrently with the event(s) being documented



Records Creation, Documentation and Management Issues

Survey Results

- 52% always or usually follow consistent records creation procedures
- yet 61% only occasionally, if ever, rely on a GIS procedures manual
- 49% always or usually follow standardized and/or documented file naming/versioning procedures



Accuracy, Reliability and Authenticity Issues

Case Study Results

- Tend to approach these issues pragmatically
- Little reliance on, or concern for, formalized procedures for addressing them
- Instead, informal assessments are based on presumptions that underlie professional authority, together with an individual's reputation for quality fieldwork within the archaeological community and published critical analyses by other experts



Accuracy, Reliability and Authenticity Issues

Case Study Results

- Establishing authenticity is less of a concern
 - Again, assumption of authenticity based primarily on presumptions that underlie professional authority
 - Hence, reasonable assumption of authenticity exists whenever data/records come from a state repository or fellow researcher, for example



Accuracy, Reliability and Authenticity Issues

Case Study Results

- Only use basic password security, no sophisticated security, access or auditing procedures used
- No formal, documented policies or procedures governing access to database/GIS records
- Access requests are assessed by GIS specialist on ad hoc, case-by-case basis
- At present, only provides locked (i.e., read-only) version of the database



Accuracy, Reliability and Authenticity Issues

Case Study Results

- GIS Specialist and volunteer are only persons with unrestricted access to the database
- Record-making, access and modification competences not formally documented in any policies or procedures manuals
- GIS Specialist Considers project too small-scale to warrant more stringent or elaborate access and security monitoring procedures



Accuracy, Reliability and Authenticity Issues

Survey Results

- Considerable variability in the way the term accuracy is defined and used
- “closeness to ground truth” or “how closely GIS data represents the real-world” most common usage (45%)
- Significantly, 33% mentioned integral role that documentation of datasets plays in assessments of accuracy



Accuracy, Reliability and Authenticity Issues

Survey Results

- With regard to authenticity, 64% always or usually take measures to ensure record creators are explicitly identified when sharing GIS projects with other researchers or general public
- With regard to reliability and accuracy:
 - 35% never formally audit their data in any way
 - 25% do so only occasionally
 - 60% use some type of formal access/security procedure to protect GIS files from unauthorized access or modification



Accuracy, Reliability and Authenticity Issues

Survey Results

- 54% absolutely or strongly confident their data have never been tampered with or become corrupted over time in such a way that would reduce their value for future use or make it impossible to reproduce or recreate their original results
- 12% expressed little or no confidence in the long term integrity of their GIS data



Appraisal Issues

Case Study Results

- No formal records appraisal procedures in place
- Instead, “appraisal” based entirely on idiosyncratic criteria administered ad hoc with little, if any, documentation
- However, underlying “practicality” rationale helps guide appraisal decisions with respect to long-term preservation of intermediary files
 - too numerous and often too large to retain and manage effectively
 - usually easier to recreate when/if needed



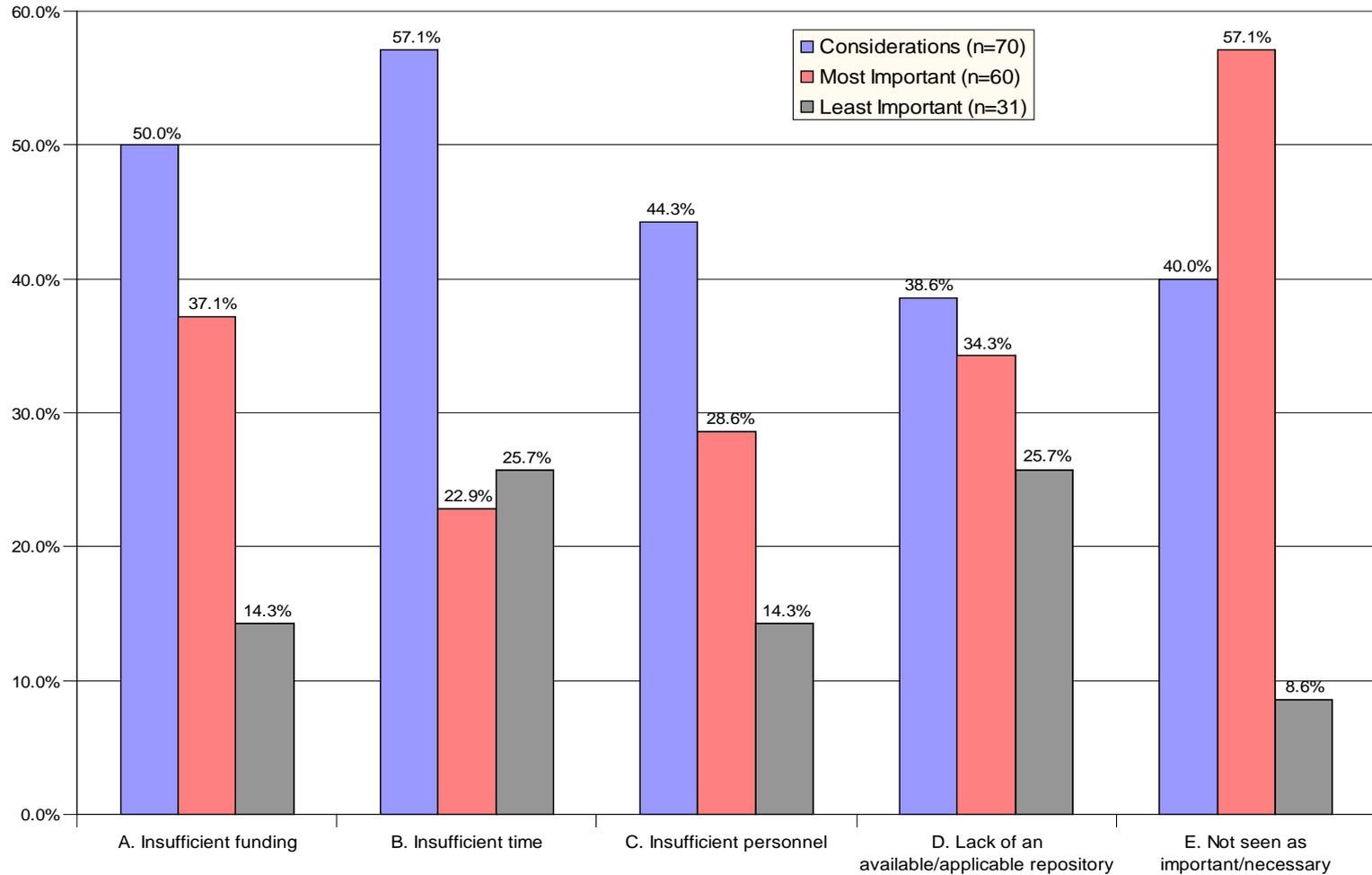
Appraisal Issues

Survey Results

- No direct questions on appraisal were asked
- Asked which factors prevented long-term preservation of some or all GIS projects
- Also asked to indicate most / least important factors
 - ‘Insufficient time’ most commonly cited (57%), but ranked last in terms of importance (23%)
 - 40% did not consider long-term preservation important or necessary
 - 57% cited this as number one reason why they don’t attempt to save their GIS projects for the long term



Appraisal Issues



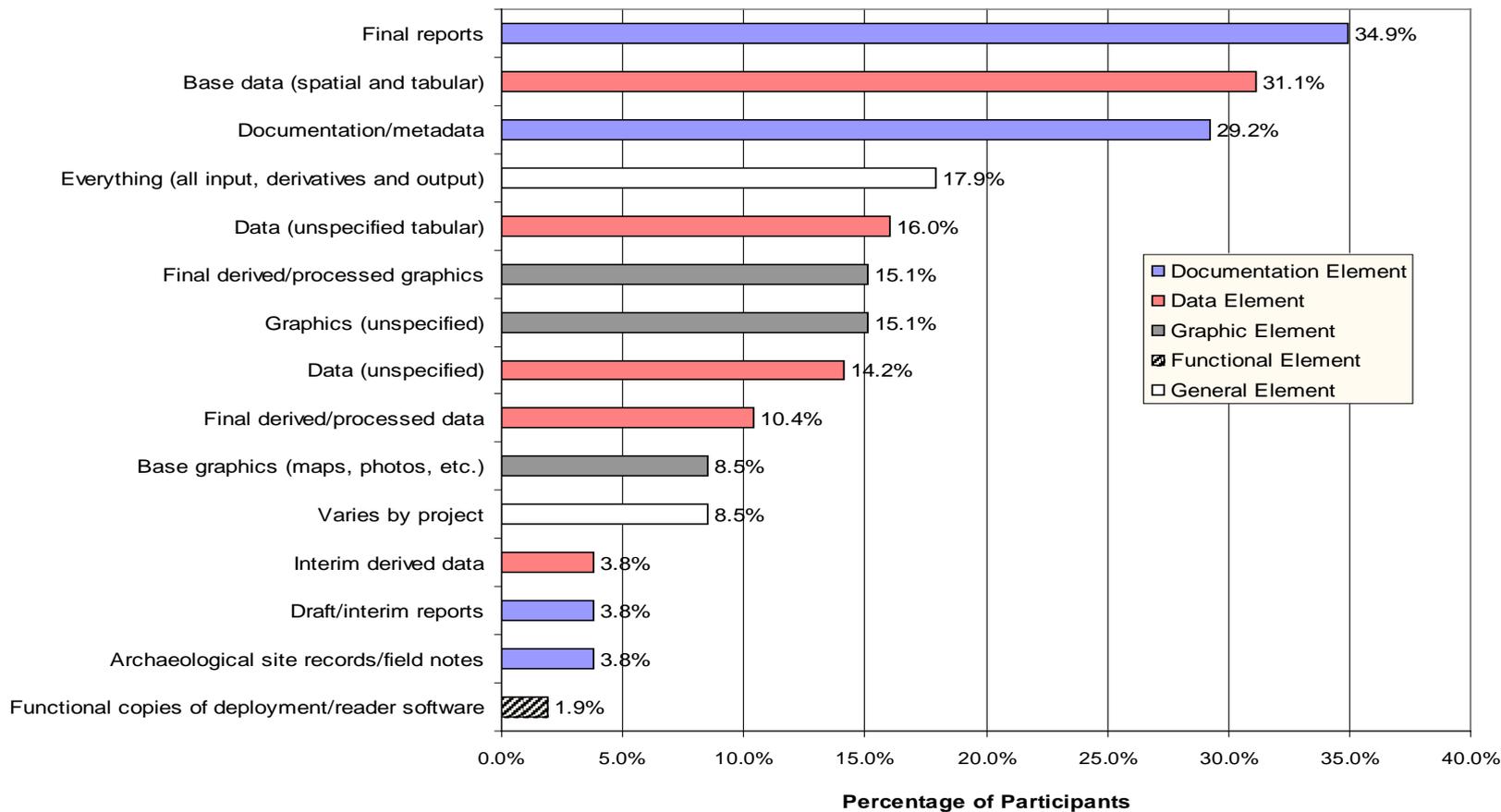
Appraisal Issues

Survey Results

- Asked to identify which elements, outputs, etc. of GIS projects should be preserved for future use/reference and why
- Participants identified 5 broad element categories:
 1. **documentation** (e.g., reports, documentation, metadata, etc.)
 2. **data** (e.g., base, derived and interpreted data)
 3. **graphic** (e.g., maps, aerial photos, etc.)
 4. **functional** (e.g., working copy of GIS software)
 5. **general** (catch-all for “save everything” to “it varies from project to project”)



GIS Elements Deemed Desirable to Preserve Long-term



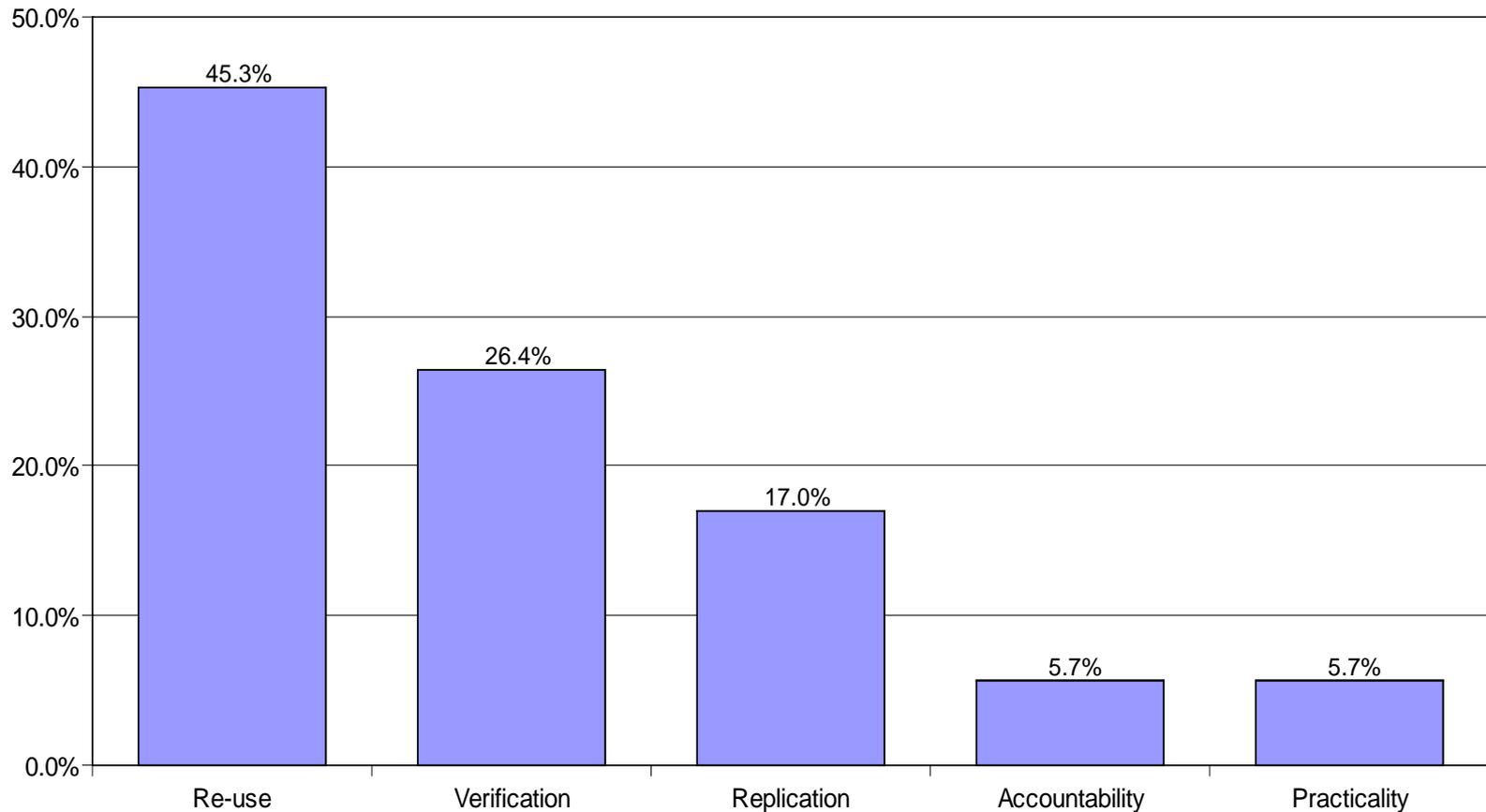
Appraisal Issues

Survey Results

- Rationales for why also fell into 5 broad categories:
 1. **Re-use** of data by other researchers
 2. **Verification** of project's analyses, methodology, etc. by other researchers
 3. **Replication** of the project by other researchers
 4. **Accountability** in case project's data or results are challenged
 5. **Practical reasons** related to long term preservation (physical, logistical, financial)



Rationales for Long-term Preservation



Long-term Preservation Issues

Case Study Results

- No formal or systematic long-term preservation procedures
- Short-term preservation strategies (i.e., backups burned to CDs) used on limited, largely ad hoc basis (before “significant” changes to database)
- Determination of “significant” is entirely ad hoc, undocumented, case-by-case decision
- Purpose of backups is to enable roll-back to earlier version should any of the new data later be found to be inaccurate or corrupt



Long-term Preservation Issues

Case Study Results

- Periodically migrate data and system files to newer versions of software, primarily to address immediate software obsolescence and current file usability issues, rather than in response to long-term preservation concerns
- No procedures to verify successful transfer of records or to monitor/assess effectiveness of preservation strategies
- No procedures to monitor continuing integrity of backups



Long-term Preservation Issues

Case Study Results

- Lack of procedures is due to:
 - Limited financial and personnel resources
 - GIS Specialist's view on importance of long-term preservation of GIS projects
 - Believes no one will need his data or results within 20 years (will be “obsolete”)
 - Notes also that researchers rarely utilize data and results of archaeologists from 100, 50 or even 20 years ago
 - Believes everything about the project worth preserving long-term is in the publications of results



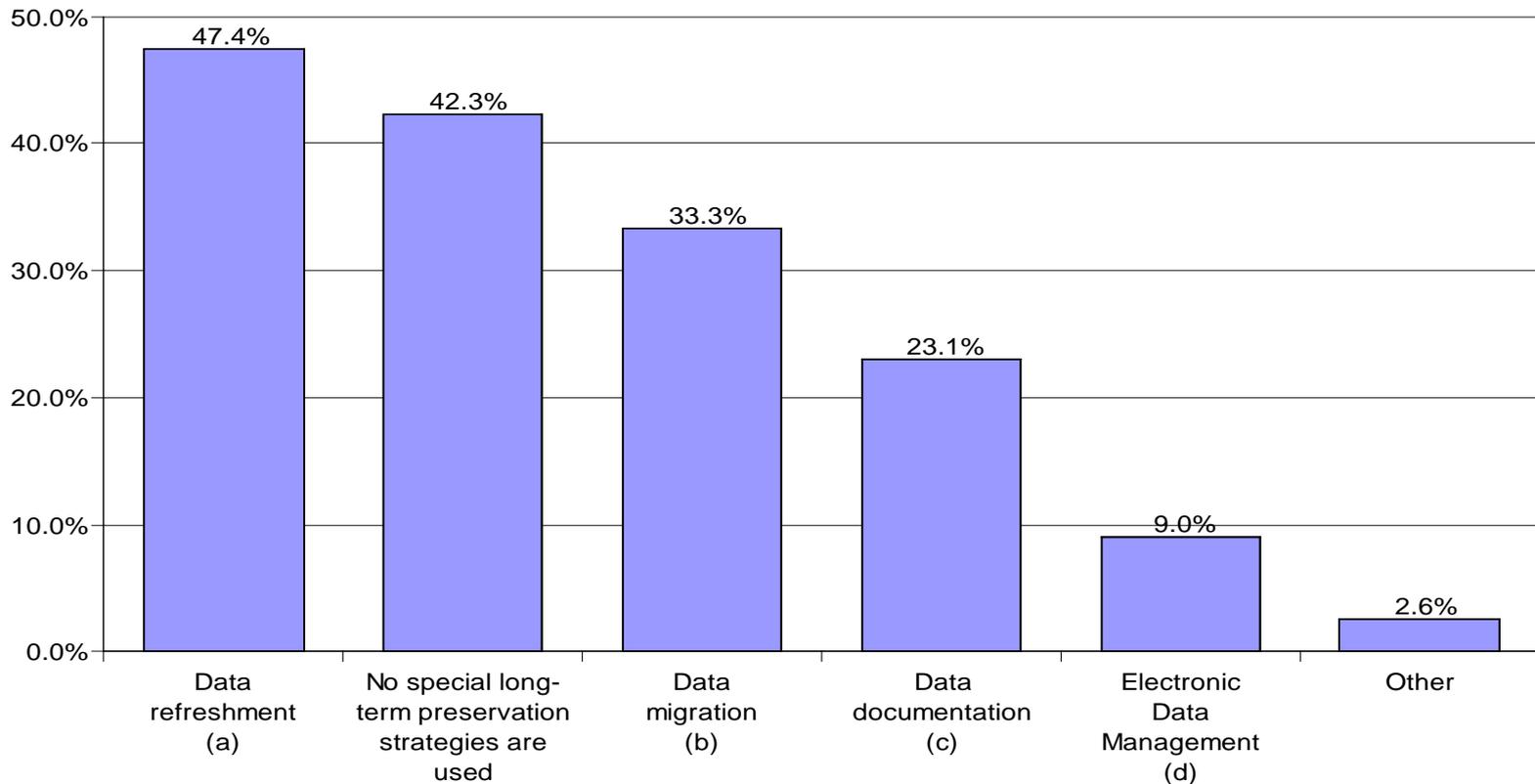
Long-term Preservation Issues

Survey Results

- 73% concerned about transfer of their GIS projects to an archives
- Yet, only 26% always or usually transfer their completed projects to an archives
- Instead, 59% rely on “in-house” long-term preservation strategies
- 53% of these said “in-house” strategies are implemented in an irregular and subjective manner without aid of any established standards or guidelines



“In-house” Preservation Strategies



(a) i.e., copying files from one medium to the next as the original medium nears the end of its reliable life span

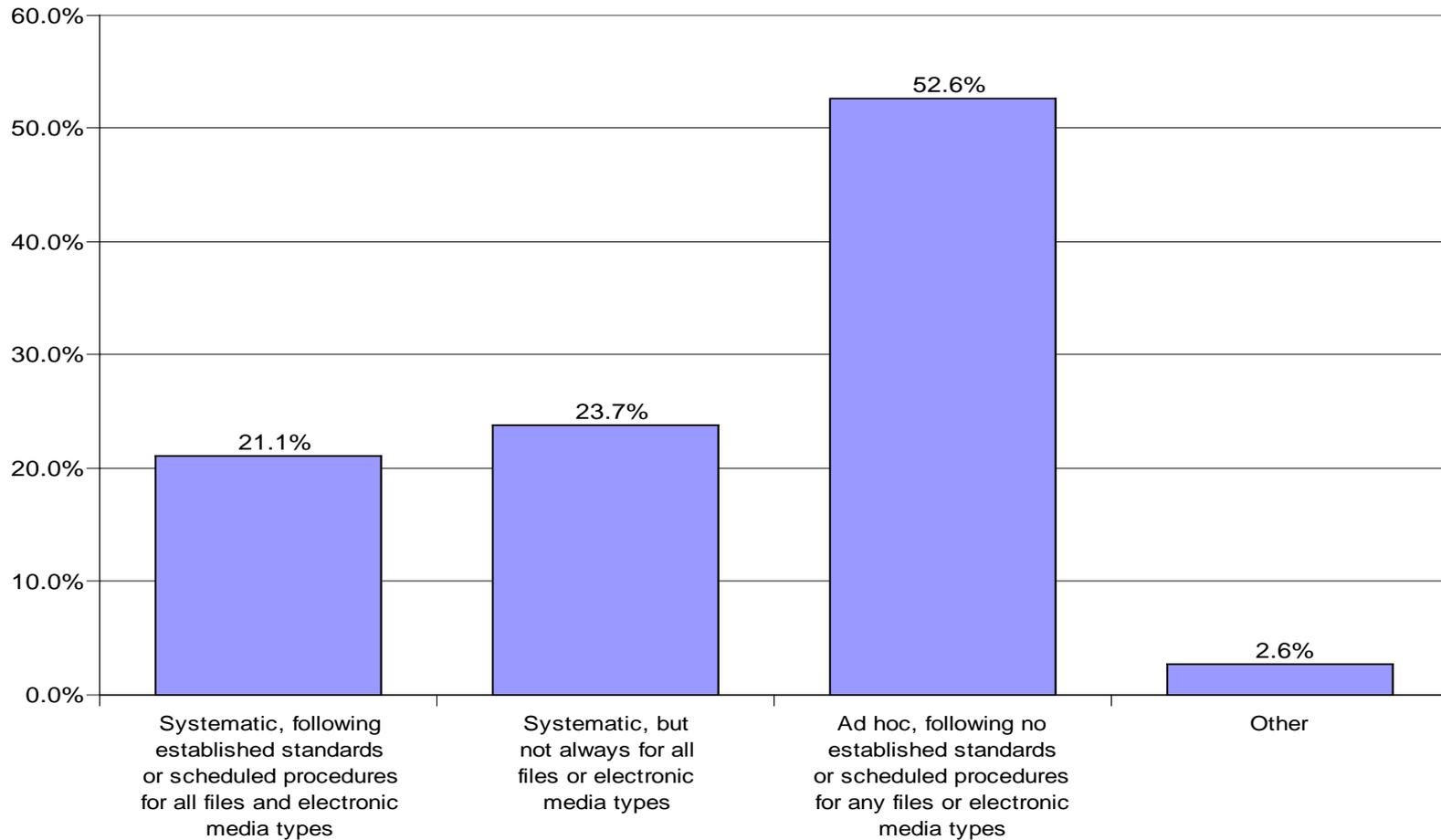
(b) i.e., converting files from one format or structure into another that can be read by current versions of software

(c) e.g., to track and explain data migrations, abbreviations used, file naming conventions, etc.

(d) i.e., data management databases that automatically indicate when files need backing up, migration or refreshment



Implementation of Strategies



Conclusions

- Discernable level of awareness of issues related to long-term preservation of authentic, accurate and reliable records
- But no concerted, profession-wide response to these issues
- Case study and survey data concur that most record-keeping procedures and practices to date are, at best, ad hoc and idiosyncratic
- Idiosyncratic approach is because a majority of GIS archaeologists have little or no formal GIS training



Conclusions

- Consequences of this 'self-taught' approach most apparent in the general disregard of formal and systematic procedures for records creation, maintenance, documentation and preservation.
- Result is a widespread, general inability within the archaeological profession of ensuring the creation, maintenance and long-term preservation of authentic, reliable and accurate digital records.

