MINNESOTA
GEOSPATIAL ADVISORY COUNCIL

Minnesota Geospatial Advisory Council Archiving Workgroup
2018 Public Geospatial Data Archive Survey

Introduction

The **Minnesota Geospatial Advisory Council** (GAC) is authorized to work as a coordinating body for the Minnesota geospatial community, representing a cross-section of organizations that include city, county, regional, state, federal and tribal governments as well as education, business and nonprofit sectors, and other stakeholder groups that benefit from geospatial technology.

Data archiving of geospatial information encompasses a wide range of considerations and practices for preserving public geospatial records and historical materials. In fall of 2018, the GAC authorized the creation of an **Archiving Workgroup** http://www.mngeo.state.mn.us/workgroup/archiving/ with the purpose of defining the policies, best practices, and procedures for archiving geospatial data in Minnesota, so that a wealth of valuable geospatial data can be preserved and available for future use. The Archiving Workgroup aims to engage with data stewards and stakeholders at various levels of government, academic institutions, private sector interests, non-profit organizations and citizens of the state, and to collaborate with MnGeo to propose datasets and methods for geospatial data archiving.

Purpose of the Survey

This survey was intended to provide a framework to capture the relevant topics of interest toward data archiving methods, prioritization of archival dataset pilot projects, and related data archiving concerns such as retention policies, licenses and data use agreements. A draft of the survey was presented in a preliminary format at the MN GIS/LIS conference in Duluth at a table display outside of the Geolounge. Feedback was solicited from as many conference attendees as possible by engaging in first-hand discussion and having the survey available for testing. There were 10 respondent that tested the preliminary draft survey and there were dozens of conversations with stakeholders that resulted in improvements to the final survey. Suggestions received that were incorporated into the final survey included recommendations for improving the scope of the survey responses and including open-ended feedback that was offered. The final survey had a greater number of response options for the priority data list, increasing the sector representation and areas of expertise offered in the responses, and response options that were open-ended for write-in responses. The final survey was distributed widely to the geospatial community in Minnesota for stakeholder input.

Survey Sample

The final version of the 2018 Public Geospatial Data Archive Survey was offered in several accessible formats for public and stakeholder feedback between 12/3/2018 and 12/21/2018. The survey link was sent to professional and public list-serves by members of the Archiving Workgroup; to the MN GIS/LIS e-

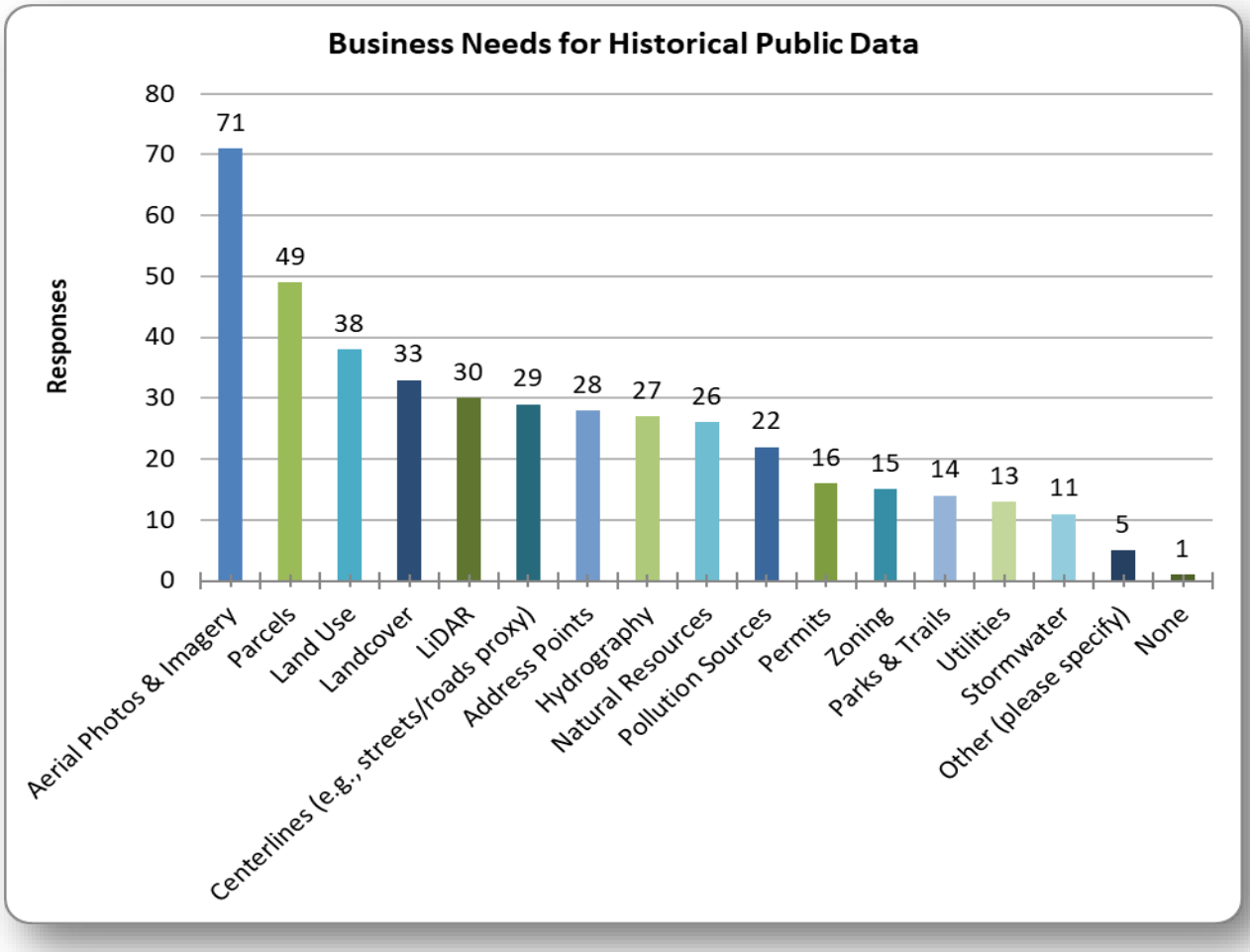
news list-serve; and was posted for general public access on the Archival Workgroup Website. There were 88 respondents to the survey in total, representing a diversity of GIS sectors and industries (as specified below in summary of questions 7 and 8). A majority of the respondents identified themselves as being associated with the transportation and environmental sectors in local and state government. The Archiving Workgroup discussed the potential for bias in the results due to this sector affiliation and agreed that the distribution of the survey respondents is comparable to the population of GIS professional stakeholders affected by this effort. As such, the survey results were not normalized by sector representation. However, the results were adjusted for weighted averages by omitting the 'Not Applicable' responses that were recorded from the weighted average calculations.

Results

Each question is summarized below, including any optional text that was written in by respondents.

Question 1:
Which historical public data does your organization have the greatest business need for preserving?
(Check all that apply)

Answered: 88
Skipped: 0



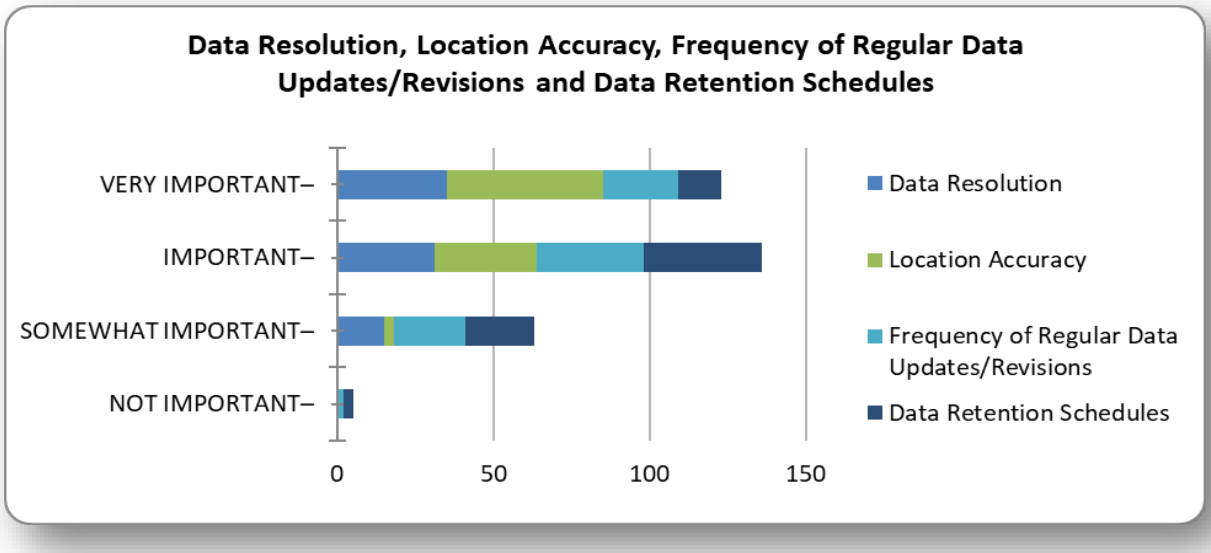
Historical aerial photos were the most frequently noted as being the highest business need for respondents, with over 80 percent of respondents marking it as an important dataset for historical reference. Secondary to that were Parcels (55%), Land Use (43%), Landcover (37%), LiDAR (34%), Centerlines (33%), Address Points (32%) and Hydrography (31%). The other datasets were least often noted as having an archival business need. Other responses included: source water protections, census tracts, section corners, and demographics.

Question 1: Answer Choices	Responses	Percentage
Aerial Photos & Imagery	71	80.68%
Parcels	49	55.68%
Land Use	38	43.18%
Landcover	33	37.50%
LiDAR	30	34.09%
Centerlines (e.g., streets/roads proxy)	29	32.95%
Address Points	28	31.81%
Hydrography	27	30.68%
Natural Resources	26	29.54%
Pollution Sources	22	25.00%
Permits	16	18.18
Zoning	15	17.05%
Parks & Trails	14	15.90%
Utilities	13	14.77%
Stormwater	11	12.50%
Other (please specify)	5	5.68%
None	1	1.14%

Question 2: How important are the following characteristics of geospatial data regarding their usefulness as an archived dataset?

Answered: 87
Skipped: 1

The usefulness of archived geospatial data was assessed based on the following characteristics: data resolution, location accuracy, frequency of regular updates/revisions, and data retention schedules. The most important of these was location accuracy, followed by data resolution, frequency of updates, and finally data retention schedules. One other response stated: "The more complete the metadata, the more useful the archived data will be."

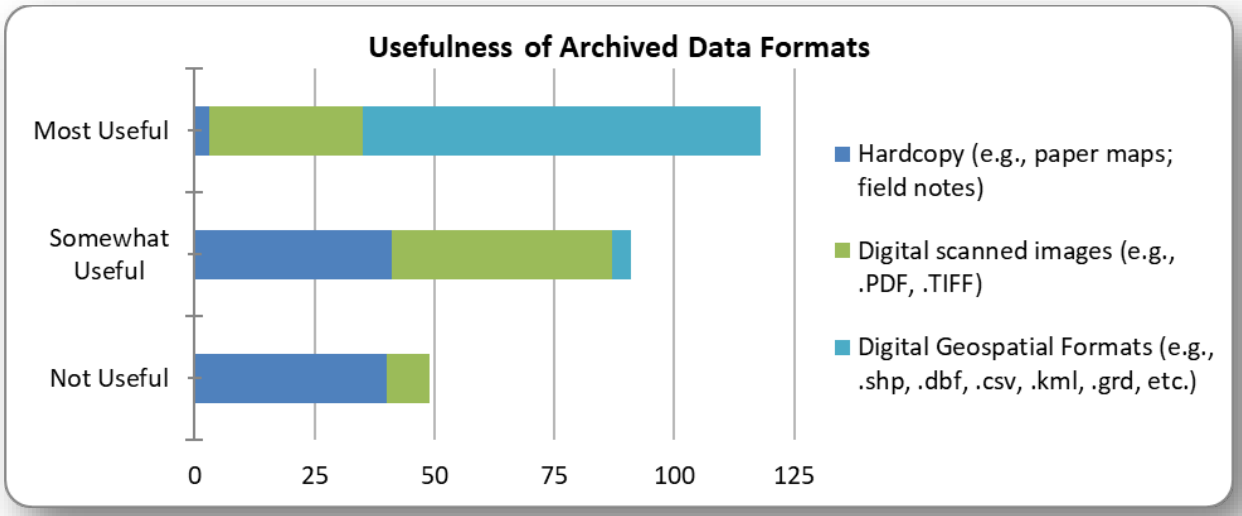


Question 2: Answer Choices	Very Important	Important	Somewhat Important	Not Important	Total	Weighted Average
Location Accuracy	50	33	3	0	86	3.55
Data Resolution	35	31	15	0	81	3.25
Frequency of Regular Data Updates/Revisions	24	34	23	2	83	2.96
Data Retention Schedules	14	38	22	3	77	2.82

Question 3: Which data format would be most useful for your sector? How useful would a data archive be for your sector based upon the following archived formats:

The usefulness of archived geospatial data was assessed based on the preferred data format. The most useful data format for geospatial archives was determined to be digital geospatial formats, including common file types such as shapefiles, database files, comma-separated values, keyhole markup language and gridded data. These digital file formats were selected as most important by 83 of 88 respondents. The digital scanned images were determined to be both most useful and somewhat useful data formats, having 32 and 46 answers for each, respectively. The least useful format for archiving purposes was hardcopy, including paper maps and field notes as examples. The hardcopy answer choice was also selected as somewhat useful by 41 responders as not useful by 40 responders and was selected as most important by 3 of 84 responders.

Other comments included: "Can't choose a most useful since it depends on the dataset; digital would be the most useful except that if formats are not migrated with changing technology, hardcopy originals may be needed since they're always readable."; "Digital Services are preferred, especially open data options. Non-proprietary formats, etc."; "digital data formats are tricky, they can change or become obsolete. CSV and DBF might be most stable."; and "Web map services or web feature services would be best."



Question 3: Answer Choices	Most Useful	Somewhat Useful	Not Useful	Total	Weighted Average
Hardcopy (e.g., paper maps; field notes)	3	41	40	84	1.56
Digital scanned images (e.g., .PDF, .TIFF)	32	46	9	87	2.26
Digital Geospatial Formats (e.g., .shp, .dbf, .csv, .kml, .grd, etc.)	83	4	0	87	2.95

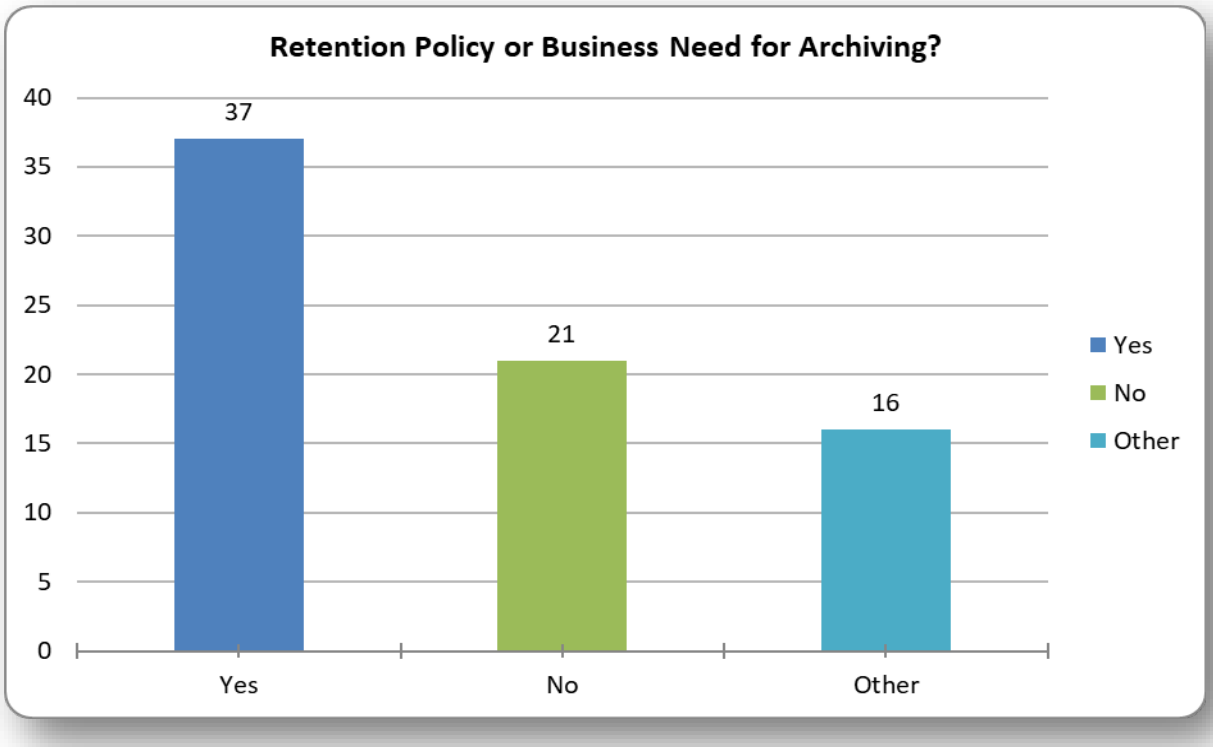
Question 4: Does your organization have a retention policy or a business need for archiving proprietary geospatial data?

Answered: 87
Skipped: 1

The significance of having archives was assessed by asking if stakeholders had a retention policy or business need for geospatial data archiving or their spatial data. The responses indicate that nearly half of the respondents (37 of 87) do have an organizational retention policy or business need for archiving geospatial data. Approximately one-quarter of respondents answered 'No' (21 of 87) and the remainder answered as 'Not Applicable' (13 of 87) or wrote in other responses (16 of 87).

The other responses are listed herein:

- 6 responses were "I don't know", "I have no idea", "Unknown", and "Not Sure"
- "A recommendation would be helpful."
- "Proprietary? We have a business need for archiving spatial data that we have produced. It is public data."
- "this area is still the Wild Wild West without standards"
- "Sort of but its based on the type of data, nothing global for all spatial data"
- "Common sense retention based on activity and change"
- "Probably, whether it is actually adhered to is a better Question."



Question 4: Answer Choices	Responses	Percentage
Yes	37	50.00%
No	21	28.38%
Other	16	21.62%

Question 5: Please classify the following considerations toward contributing data to a geospatial archive:

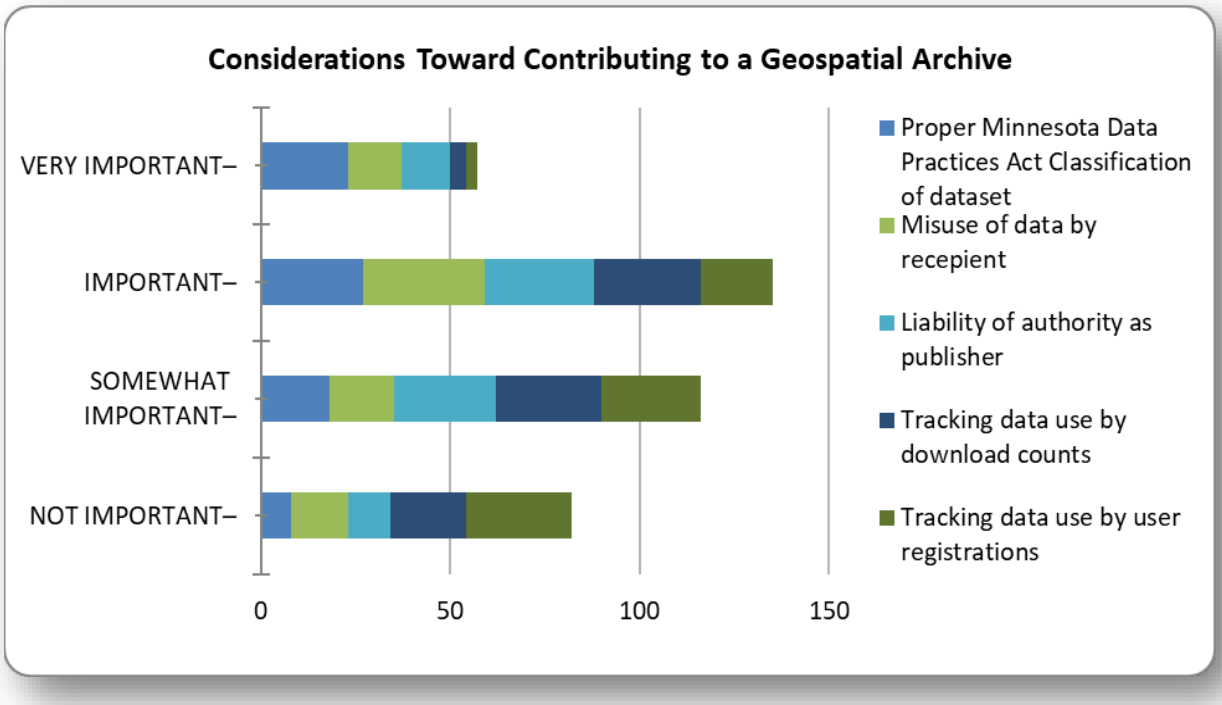
Answered: 85
Skipped: 3

Survey responders were asked to rank their considerations toward contributing data to a geospatial archive for access by users. Most respondents ranked all of the considerations as very important, important, or somewhat important, including: proper Minnesota data practices act classification of dataset (68 of 85), misuse of data by recipient (63 of 85), liability of authority as publisher (69 of 85) and tracking data use by download counts or user registrations (60 and 48 of 85, respectively). The consideration that was mot frequently noted as being most important was the proper Minnesota data practices act classification of the dataset, followed by misuse of the data, liability as publisher and tracking use by downloads or registrations.

Other write-in responses included the following:

- "We have no control over how recipients use data - provide metadata to support informed use. Not in favor of barriers to access such as user registration."

- "Some data has tremendous value for a limited number of people, I would not want to kill a dataset because it has few downloads, or even zero."
- "Mostly interested in how the data is reused. Best to make the data very easy to get at vs encouraging the reuse of data."



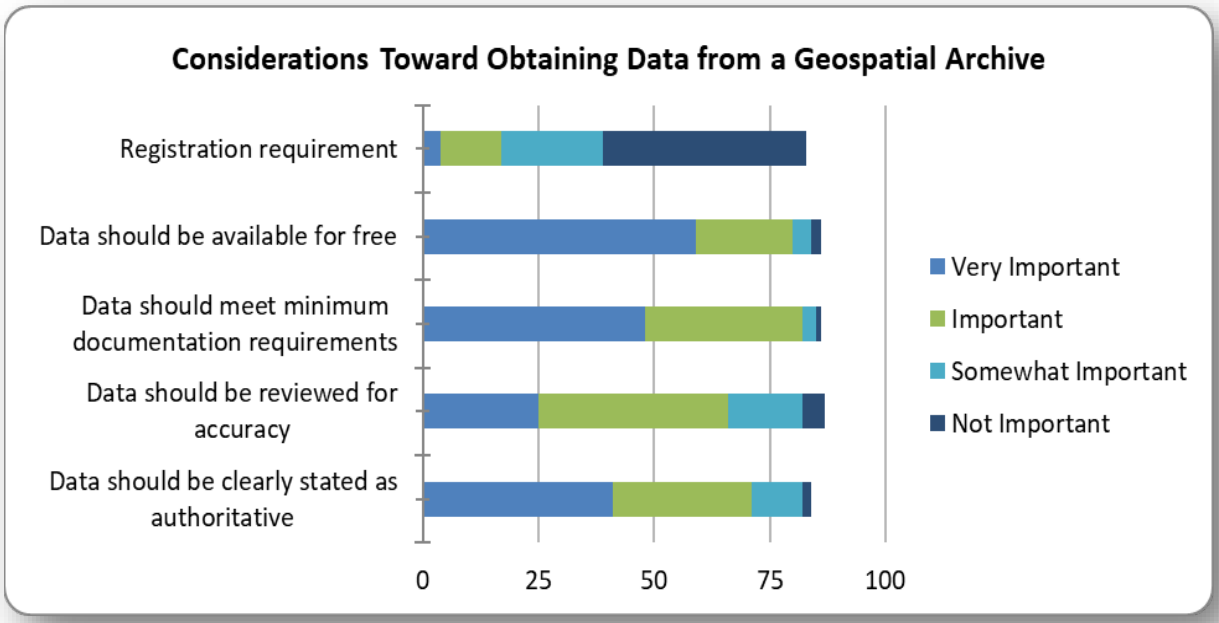
Question 5: Answer Choices	Very Important	Important	Somewhat Important	Not Important	Total	Weighted Average
Proper Minnesota Data Practices Act Classification of dataset	23	27	18	8	76	2.86
Misuse of data by recipient	14	32	17	15	78	2.58
Liability of authority as publisher	13	29	27	11	80	2.55
Tracking data use by download counts	4	28	28	20	80	2.2
Tracking data use by user registrations	3	19	26	28	76	1.96

Question 6: Please classify the following considerations toward obtaining data from a geospatial archive:

Answered: 87
Skipped: 1

Survey respondents were asked to rank the importance of their considerations toward obtaining data from a geospatial archive repository. The choices were all considered important by most

respondents, except for the registration requirement which was considered as 'not important' by 44 of 84 respondents. The most commonly noted important considerations toward obtaining data from a geospatial archive included both that data should be available for free (84 of 87) and that the minimum documentation requirements (85 of 87). It was also determined that is important for data to be clearly stated as authoritative and have a time stamp of currency associated with it (82 of 87 respondents chose either very important, important, or somewhat important).



Question 6: Answer Choices	Very Important	Important	Somewhat Important	Not Important	Total	Weighted Average
Data should be available for free	59	21	4	2	86	3.59
Data should meet minimum documentation requirements (i.e., authorship & contact information)	48	34	3	1	86	3.5
Data should be clearly stated as authoritative with currency information	41	30	11	2	84	3.31
Data should be reviewed for accuracy	25	41	16	5	87	2.99
Registration requirement	4	13	22	44	83	1.72

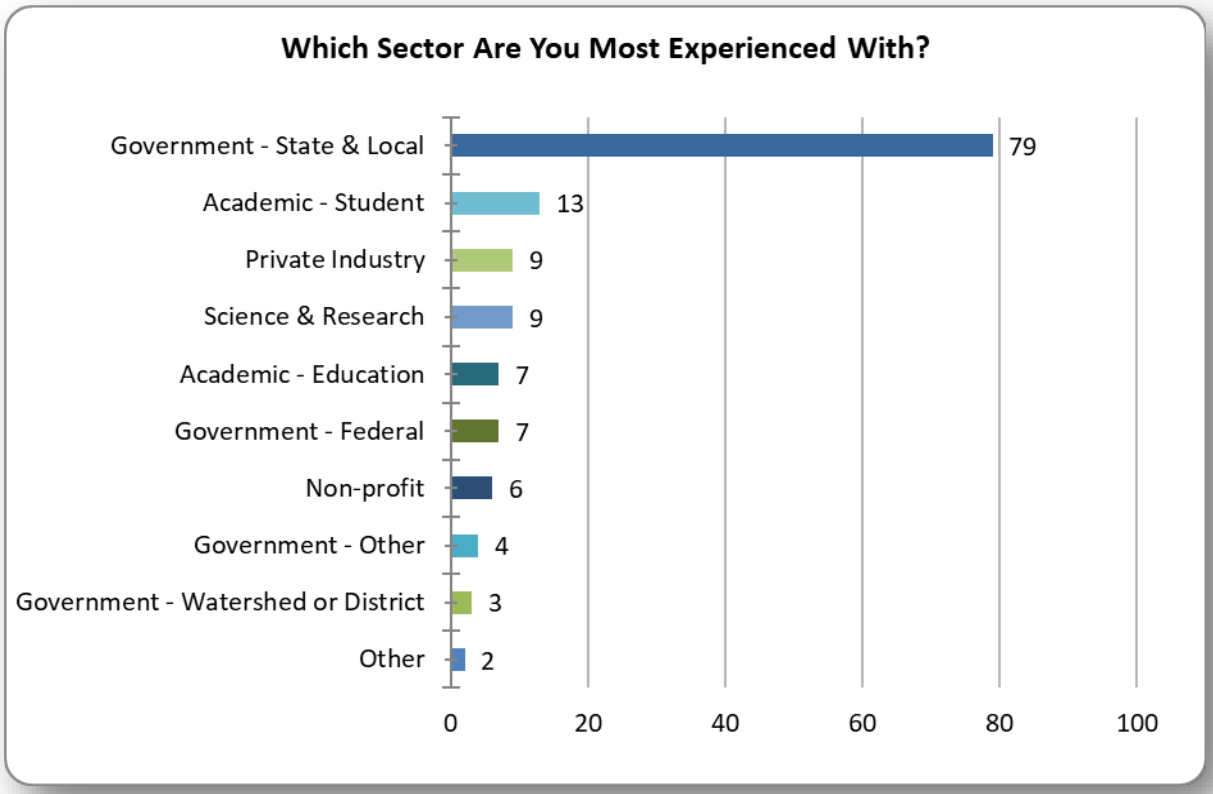
Other responses to question 6 are listed herein:

- "Some cursory checks such as making sure data is what it is labeled, digital formats open; submitting authority should be responsible for accuracy checks."
- "Basic metadata needed, but a full-on Standard metadata record may not be required for most datasets."
- "I need to know how much confidence to put in the data (which I can get from the metadata), but the spatial data doesn't have to be perfect."

Question 7: Which sector of the GIS community are you (most) experienced with?

Answered: 88
Skipped: 0

The majority of the survey respondents were most experienced working with state and local government (79 of 88 individual responders). This question was designed to allow more than one answer per respondent and in total there were 139 sector associations noted by 88 individuals. On average, each respondent identified 1.6 sectors that were most related to their GIS experience. Other responses included the following responses: "personal use" and "Tribal".



Question 7: Answer Choices	Responses	Percentage
Government - State & Local	79	89.77%
Academic - Student	13	14.77%
Science & Research	9	10.23%
Private Industry	9	10.23%
Government - Federal	7	7.95%
Academic - Education	7	7.95%
Non-profit	6	6.82%
Government - Other	4	4.55%
Government - Watershed or District	3	3.41%
Other	2	2.27%

Question 8: What is your GIS industry and/or area of expertise?

Answered: 88
Skipped: 0

Respondents were asked to identify which GIS industry or area of expertise they identify with. Overall there were 216 areas of expertise chosen by 88 respondents, resulting in an average of 2.5 industry associations per individual. The most popular area of expertise was associated with the transportation industry (36 of 88 individual responders). Next most popular were the environmental, natural resources, and water related areas of expertise (29 and 24 of 88 respondents, respectively).

Other responses were recorded that included: local government, real estate, GIS coordination and consulting on statewide level, public safety, taxes, revenue, demographics, historic sites and sites of significance to specific cultures.

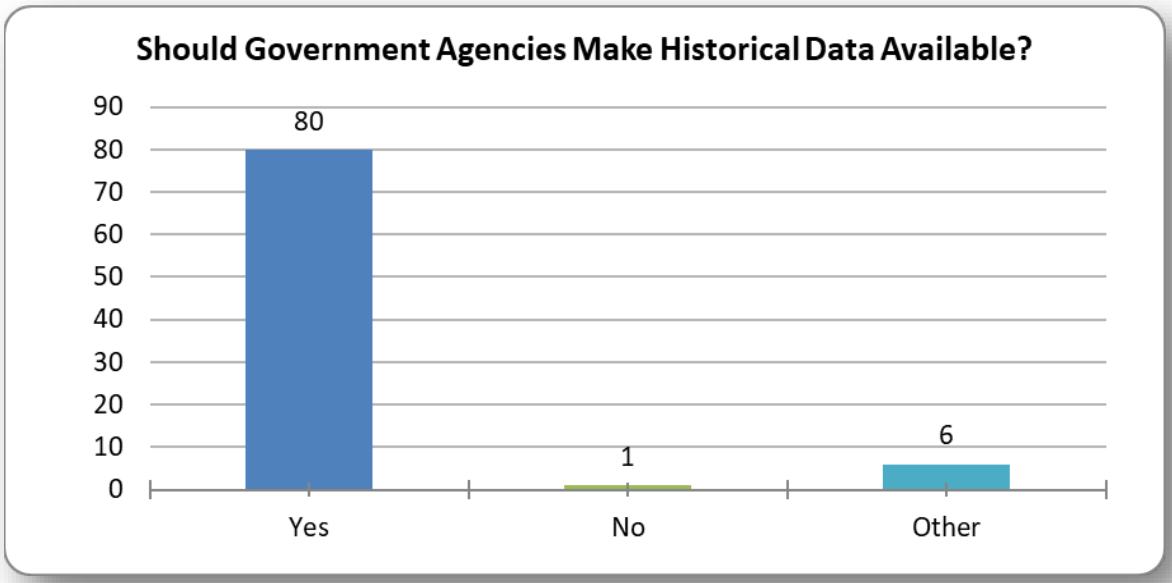
Question 8: Answer Choices	Responses	Percentage
Transportation	36	40.91%
Environmental	29	32.95%
Natural Resources	24	27.27%
Water	23	26.14%
Land Use Planning	17	19.32%
Public Health	17	19.32%
Assessment & Land Records	16	18.18%
Engineering	12	13.64%
Emergency Response	11	12.50%
Public Works/Utilities/Asset Management	10	11.36%
Other	8	9.09%
Education	7	7.95%
Marketing & Real Estate	2	2.27%
Telecommunications	2	2.27%
Banking & Retail	1	1.14%
Insurance	1	1.14%

Question 9: Do you think that government agencies should make their historical geospatial data available?

Answered: 87
Skipped: 1

The final survey feedback question focused on whether government agencies historical data should be made available. The majority of responses were 'yes' (80 of 87 respondents). One individual chose the answer 'No'. There were 6 respondents that selected 'Other' and several that wrote individual responses. The other responses are listed herein:

- "It depends on the context."
- "unsure."
- "depends on the data type"
- "If applicable. If data is meaningful for trend analysis, yes. if not, no."
- "Yes, however, if it fits their business need. Each authoritative source should make that decision for themselves."
- "Depends on the data set. Some would be ridiculous to keep historical records for, so each needs a cost/benefit analysis before deciding."



Question 9: Answer Choices	Responses	Percentage
Yes	80	91.95%
No	1	1.15%
Other	6	6.90%

Question 10: Please share your contact information if you would like to be contacted for further updates from the GAC Archival Workgroup.

Answered: 17
Skipped: 71

Respondents' personal information is being kept confidential and is used for direct communication purposes only.

Report compiled by Andra Mathews, GISP
February 7, 2019
On behalf of the Geospatial Advisory Council, Archiving Workgroup

For more information please contact the Archiving Workgroup Chair:

Ryan Mattke
Map & Geospatial Information Librarian
Head, John R. Borchert Map Library
Adjunct Faculty, MGIS Program
Project Lead, Big Ten Academic Alliance Geospatial Data Project
Co-Director, Mapping Prejudice Project
Phone: 612.624.5757
Web: http://www.lib.umn.edu/borchert
ORCID: http://orcid.org/0000-0001-8816-9289