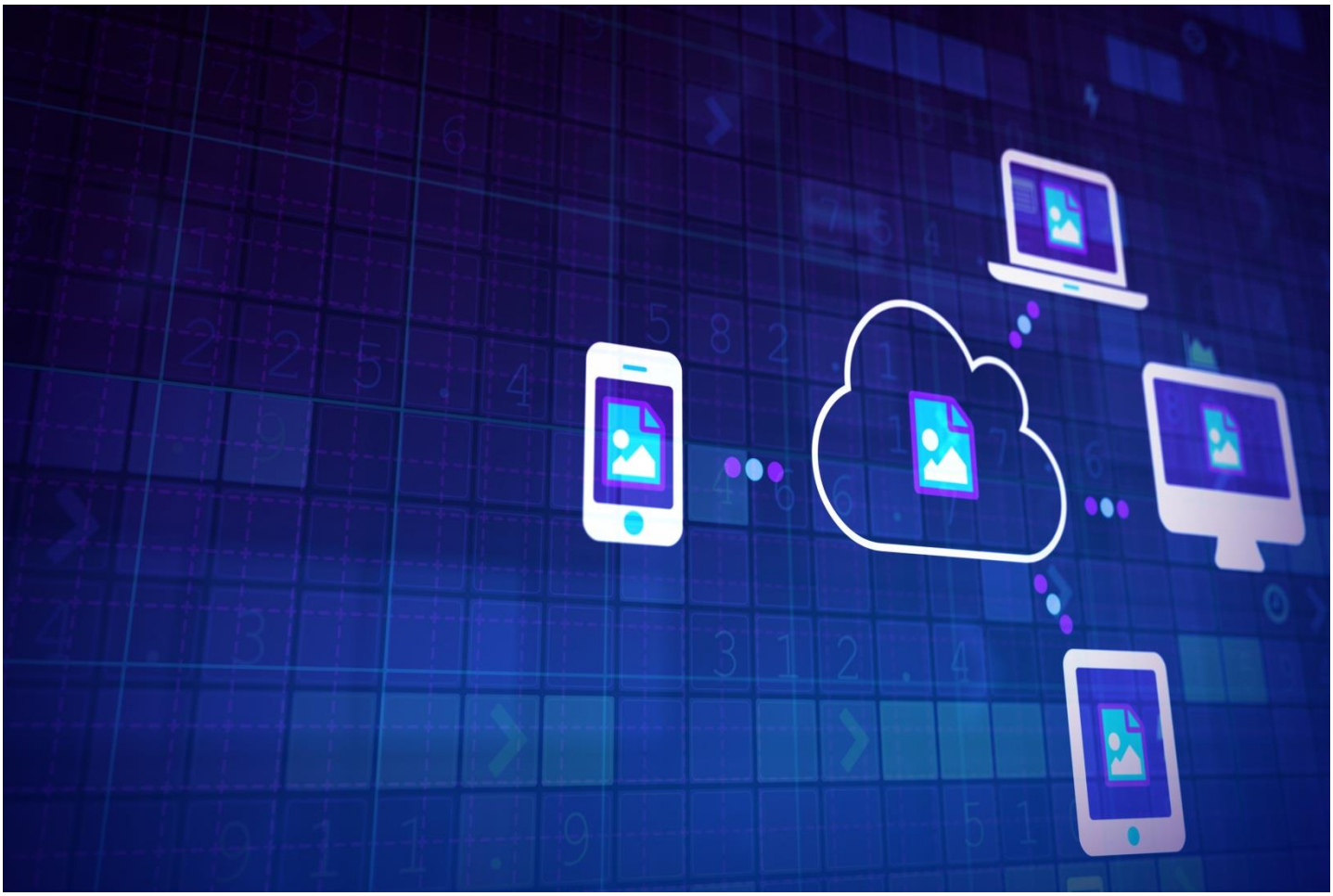
# Project Report

## Digital Records Survey 2016



Public Record Office Victoria in conjunction with Kathy Sinclair, Policy Consultant

Public Record Office Victoria

VICTORIA State Government

# Contents



**List of Acronyms**

| | |
|---|---|
| EDRMS | Electronic Document and Records Management System |
| HR | Human Resources |
| PROV | Public Record Office Victoria |
| RDA | Retention and Disposal Authority |
| TB | Terabyte (data storage size) |
| VEO | VERS Encapsulated Object (digital format required by PROV) |
| VERS | Victorian Electronic Records Strategy |

# List of Figures

# 1 Executive Summary

The Digital Records Survey Project was undertaken to obtain data from a broad selection of Victorian Government agencies regarding their:

- Current practice in the creation, management, storage and disposal of digital records; and
- Plans for the future, particularly in respect of intentions to transfer permanent public records in digital form to the state archive, Public Record Office Victoria (PROV).

The Project aimed to provide PROV with indicative data from which predictions could be made about the volume, growth, nature, quality, and management of public records in digital form. Although this objective was not fully achieved, a substantial amount of qualitative information was received which may assist PROV in future strategic planning.

Responses were elicited by use of a survey from government agencies across a range of public sector functions and are organised by theme in this report. Twenty-five agencies that commenced the survey provided responses containing useful discursive and qualitative information. The lack of quantitative data from respondents indicates poor or low level reporting measures, tools and systems.

Seven key findings emerged from this Survey. These are:

1. Half of the respondents were not confident in statistics provided for the overall volume of their digital data or estimated growth over time. Possible reasons for this include:
    a. Dispersal of data across multiple platforms and systems, including some not within the direct control of the agency (e.g. shared services);
    b. Difficulty in screening out duplication, system / application data, and error data; and
    c. Lack of information technology resources able or willing to be deployed to collection this information, which may be seen as being of limited utility for the agency itself (especially as the amount of data in storage is not usually a metric that is collated).

2. Numbers provided indicate a growth rate for digital data that is significantly high, which may be due to increased focus on information being created and captured digitally as well as difficulty in screening out duplication, system / application data and error data.

3. An understanding of what data is held, its retention requirements, and its management needs, varies across the sector. This is implied by the differences in the respondents' ability to describe their data (including how it is managed). The difference may be explained by the level of relevant knowledge and skills held combined with a trend for data to be managed by information technology rather than records and information management specialists.

4. Agencies, while committed to continuing existing local storage models, are also showing movement towards cloud-based storage solutions, with comments provided suggesting that the cost of storage was the main factor considered when determining digital solutions.

5. Ability to generate VEOs is (based on this sample) relatively low, and is outstripped by interest in executing digital transfers to PROV. This suggests agencies may need external assistance to prepare digital records for transfer in VEO format.

6. Agencies are at very different stages in their ability to transition from hard copy to digital recordkeeping. This may be due to belief that physically signed documentation must be kept for authentication purposes as two respondents indicated statutory requirements they believed mandated the keeping of hardcopy files.

7. While agencies have different focal points, key concerns in the digital records area are:
   - data security;
   - staff and system capacity;
   - costs and resourcing constraints; and
   - legacy systems and data.

Further investigations are required to expand on the findings of this project regarding the readiness of agencies to move from hard-copy recordkeeping to digital records, as well as identifying the most common shared issues across the sector for future digital data management.

# 2 Recommendations

The Digital Survey 2016 is the latest of a number of investigations by PROV into the management of digital public records in Victoria. To address the findings from the survey, three recommendations are offered. They are that PROV:

- Undertake research on how agency enterprise-wide oversight of information assets can be improved;
- Continue to develop resources to assist agencies in knowledge, competency and skill development regarding recordkeeping; and that PROV
- Continue to advocate for improved records management.

1. **Undertake research on how agency enterprise-wide oversight of information assets can be improved across Government.**

The findings (1, 2 and 3 above) show that agencies generally do not have sufficient awareness of their information assets or how to manage them effectively. PROV cannot assist agencies with the management of records if the agency cannot first identify them. PROV's new Digital Archive is designed to scale to meet future needs, but its value will depend upon identification of digital records of permanent value for preservation. A combination of high growth rates and poorly identified information assets will put considerable additional costs on the State in the delivery of public services and in preservation of the public record.

The following key topics and issues for further research and investigation were identified by the agencies surveyed: cloud technologies and storage; lack of resourcing; staff capability that is not adequate to the emerging digital world; and data security. Ongoing refinement and conduct of surveys will provide valuable and increasingly reliable longitudinal data for planning and can help ensure that PROV resources more efficiently address agency needs.

2. **Continue developing resources to assist agencies in knowledge, competency and skill development regarding recordkeeping and digital information management.**

While PROV already develops resources to assist agencies with implementing a program of records management, findings (4, 5, 6 and 7 above) indicate that topics related to the move towards digital recordkeeping should be prioritised. This includes resources related to the ability to transfer digital records in appropriate formats to PROV as well as addressing legacy systems and data.

Most respondents identified that the level of staff capability regarding the digital information management was inadequate, and indicated that costs and resourcing issues had a significant impact on this. An area for PROV to focus on is support for capability development among agency staff regarding effective management of public records and include related information sectors in order to effectively communicate digital records messages to broader agency staff.

3. **Continue to advocate for improved records management.**

The findings (4 and 6 above) indicate that agencies are in different stages of transitioning towards a fully digital future. If the government's Digital First strategic direction is to have traction,[1] then support and resources to help agencies successfully transition away from paper-based recordkeeping are required.

PROV is in a position to advocate for improved recordkeeping practices with the other government bodies involved (such as Enterprise Solutions within DPC) to ensure the Digital First implementation is successful. As well as to advocate for inclusions in information systems procurement and contract documents that address recordkeeping requirements.

---

[1] Department of Premier and Cabinet, Information Technology Strategy: Victorian Government: 2016-2020; Department of Premier and Cabinet 2017; http://www.enterprisesolutions.vic.gov.au/it-strategy/

# 3 Project Background and Goals

The Digital Records Survey Project is a successor project to an earlier interview-based research exercise, known as the Digital Data Use in Victorian Government Agencies.

The Digital Data Use in Victorian Government Agencies ran from March to July 2016. It addressed ten Victorian government agencies from various sectors and functions, via in-depth interviews and site visits. Its purpose was to provide a snapshot of agency digital data creation and use, which PROV could use to plan future strategies for managing digital records. A Final Report[2] of that project was produced and is available for reference.

The Digital Records Survey Project sought to expand on the ground covered by the earlier project through increasing the number of agencies targeted, and narrowing the scope and reach of the questions so that they were more likely to provide statistical information needed for strategic planning.

The Digital Records Survey 2016-17 project of:

- Conducting a survey of agencies from different sectors and of different functions across Victorian government;
- Using a consistent set of questions that can be re-used over time to enable benchmarking;
- Obtaining information about agency data quality and change over time. Goal information included:
  - Defensible estimates of the volume and growth year-on-year of public records in electronic format; and
  - Indicative measures of the number and type of permanent public records in electronic format.
- Analysing the results to produce a report of the 2016 survey;
- Communicating the findings of the survey to all relevant stakeholders; and
- Reviewing the project, including the Survey approach.

[2] See J:\PROV\APPRAISAL & DISPOSAL REG\Records Survey\Digital records survey (2016-0070)\Reports\PROV Digital Survey Report July 2016 KSinclair.docx

# 4 Project Methodology and Questions

The Digital Records Survey was completed by a team comprising representatives from the Standards and Policy group within PROV and an external Project Consultant, Kathy Sinclair.

Survey questions were developed collaboratively within the team and were submitted to review within PROV for refinement. A publicly accessible online survey tool was selected, the survey was designed and built, and then subjected to beta testing both within PROV and by partner agencies before deployment.

The survey was conducted via a customised form. Explanatory text, along with a link to the survey, was sent to addressees by the Project Consultant using a bcc function following an initial notification / introduction email from the Assistant Director, Government Services (PROV).

No technical difficulties were noted by any respondent in using the online survey interface, and the portal remained stable throughout the survey period. Technical issues cannot be ruled out, however, as we have no explanation at this time for those agencies who commenced the survey but did not provide any responses to the questions.

## Sample and response rate

The Survey was sent to a total of 126 agencies, compiled by the Standards and Policy team at PROV from existing contact lists. These select agencies represent the breadth of the Victorian public sector. Thirty-three agencies commenced the survey, however, only 25 provided responses to the survey questions.

Initially the survey request was sent to 53 addresses on 27 October 2016, with a second round sent to a further 73 addresses three weeks later (14 November 2016) when the set target of 30 responses was not received from the first round. From these lists there were a number of errors and messages declining the survey (30 across both rounds). Approximately 60% of the successfully delivered surveys did not elicit a response, despite a follow-up reminder email being sent two weeks after the initial request. However, the follow-up emails did provide additional responses from each list.

## Survey questions

The questions asked on the Survey were as follows:

*Section One: Data Quantities and Types*

1. What is the name of your agency / organisation?
2. At what email address would you like to be contacted?
3. What was your agency's total terabyte (TB) of stored data for:
   2012/13?
   2013/14?
   2014/15? (Numeric answer fields only)
4. Please indicate your degree of confidence in the previous answer. (Sliding scale response)
5. Please specify what storage platforms were included when determining the total TB of stored data when answering the previous question (e.g. the EDRMS, Lotus Notes, or Shared Drives).
6. Do you have a list of your agency's systems (e.g. an Information Asset Register)? (Yes / No / Other response)

*Section Two: High Value / High Risk Data*

7. Have you identified high value and/or high risk data? (Yes / No / Other response)
8. If you answered Yes to the question above, how did you identify high-value and / or high risk data?

*Section Three: System Redundancy*

9. Do you have systems that will no longer be supported or are about to be decommissioned but which contain data that is of continuing value to your agency? (Yes / No / Other response)
   **(The following three questions only appear if a "Yes" response is given to Q9)**
10. If you answered Yes to the previous question, please select all responses below that apply. Drop-down list options:
    The data will continue to be needed for access by agency staff; and/or
    The data will be needed for access by other organisations or agencies.
11. Do you have a plan to manage access to the data? (Yes / No / Other response)
12. How long will you require access to the data?
    **(All respondents see questions below this point)**

*Section Four: Data Storage*

13. What strategies are planned for the next five (5) years regarding where data will be stored? Please select all that apply:
    Continuing with the current strategy;
    A cloud storage solution;
    A shared storage solution (with other agencies); and/or
    A commercial storage solution

*Section Five: Transfer*

14. To what extent is your organisation reliant on hardcopy records to operate?
15. Are your organisation's business processes covered by a current and relevant Retention and Disposal Authority (RDA)?
16. Do you know which of your organisation's digital records or data are required for ten (10) years or longer (for example, which are classified by an RDA as being of permanent retention)?
    Drop-down list options:
    Yes for all systems used;
    Yes for the recordkeeping system/s only;
    Yes for some systems but not all;
    No; or
    I don't know.
17. What is your capacity to create / generate VEOs (VERS Encapsulated Objects)?
18. Do you intend to transfer digital records to PROV within the next five (5) years? (Yes / No / Other response)
19. What barriers to transferring digital records / data do you have?
    **(The final question is an open field response)**

*Section Six: Key Issues*

20. What are your top three (3) hot topics regarding data, digital records or information management and why?

# 5 Question Responses

## Section One: Data Quantities and Types

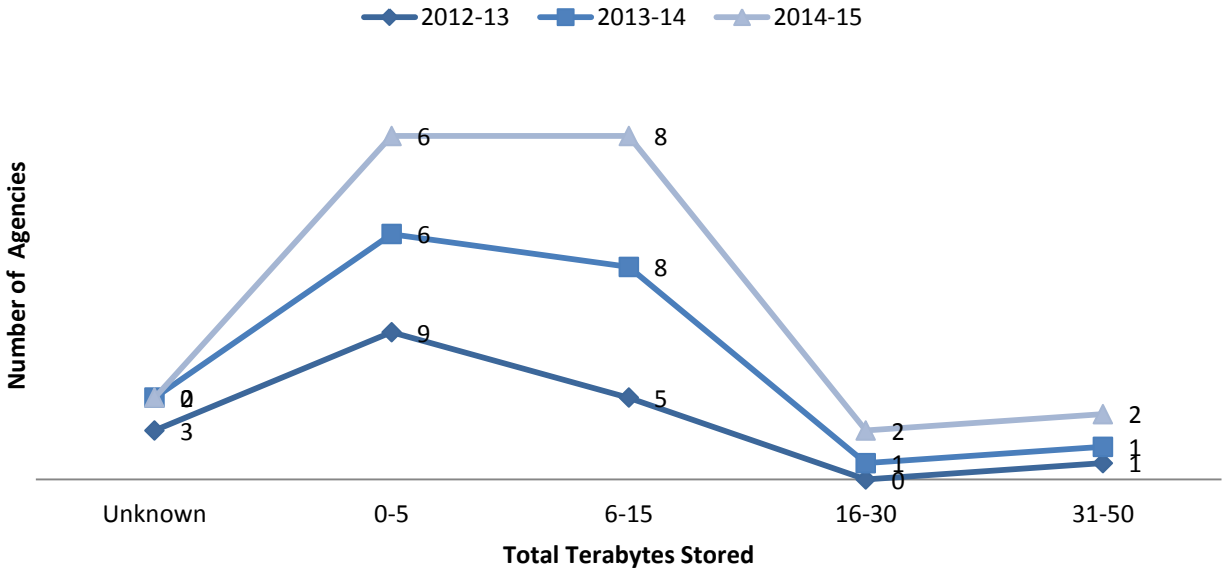
The first set of questions (Q1 through to Q6) addressed the data quantities and types that agencies had in storage.

### Key Findings

The proportion of agencies able to provide data storage estimates was considerably better than in the in-depth interview-based project, with 72% of respondents (18 of 25) completing Q3 and Q4.

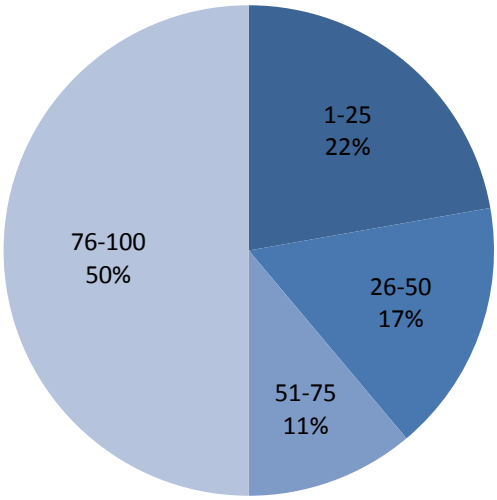
Only eleven of the 18, however, expressed greater than 50% confidence in their figures, so caution is advised when using this as a basis for extrapolation across Victorian agencies (see Figure A, below).

**Figure A: Total Terabytes Stored**




Most respondents who did answer the question regarding data levels expressed relatively modest to low levels of confidence in their responses, with only one agency that provided a full response to data storage numbers expressing complete (100%) confidence in their answer (see Figure B, below). This makes it difficult to accurately interpret the data.

## Figure B: Agency Confidence Levels

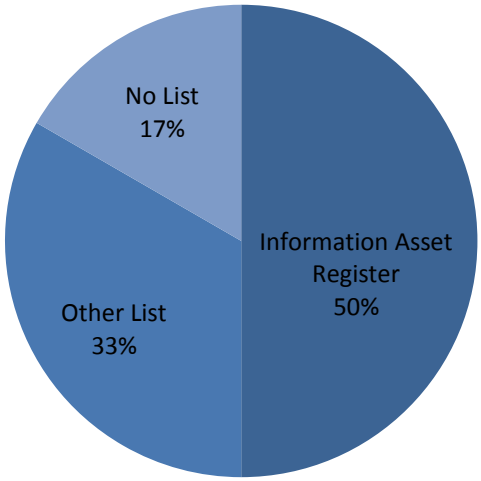



On the other hand, agencies were able to be more responsive regarding the qualitative questions regarding data systems and registers. This is also consistent with the outcomes in the earlier detailed project, where all ten agencies interviewed were able to provide narrative descriptions of their data and information management landscape and provide very detailed lists of systems in use for the creation and management of data.

In terms of systems that create and manage data, respondents overwhelmingly identified EDRMS, shared drives, mail systems, and line of business systems, including purpose-built systems in the areas of finance, human resources, payroll, and case management. SharePoint in particular is frequently cited as an overarching system (see Figure C, below). Agencies have good awareness of what systems they have with the results suggesting that 50% have Information Asset Registers with a further 33% having a list of systems (see Figure D, below).

| Figure C: Storage Platforms |
| --- |
| EDRMS / ECM<br>File Share / Shared Drives / File Servers<br>Database Storage (including SharePoint)<br>Purpose built systems<br>Email (e.g. Lotus Notes; Exchange)<br>Web applications<br>Image Database<br>Finance Systems (including Human Resource Systems / Payroll)<br>SQL Databases<br>Case Management Systems |

## Figure D: No. of Agencies with a List of Information Systems




A reasonable conclusion, based on the consistency of the results in this area across the two projects and the unavailability of data estimate figures in other government reports, is as follows: People were not confident in the numbers they provided regarding being able to accurately assess the overall volume of their digital data or estimate growth over time. Some indicated that they were on top of things where as others were unable to provide the information required. Based on the responses given in interviews in the earlier project as well as feedback given by survey respondents, it is likely that this is due to a number of factors, including:

- Dispersal of data across multiple platforms and systems, including some not within the direct control of the agency (e.g. shared services);
- Difficulty in screening out duplication, system / application data, and error data; and
- Lack of information technology resources able or willing to be deployed to collect this information, which may be seen as being of limited utility for the agency itself.

Agencies have good awareness of what systems they have with the results suggesting that 50% have Information Asset Registers with a further 33% having a list of systems. In terms of systems that create and manage data, respondents overwhelmingly identified EDRMS, shared drives, mail systems, and line of business systems, including purpose-built systems in the areas of finance, human resources, payroll, and case management.

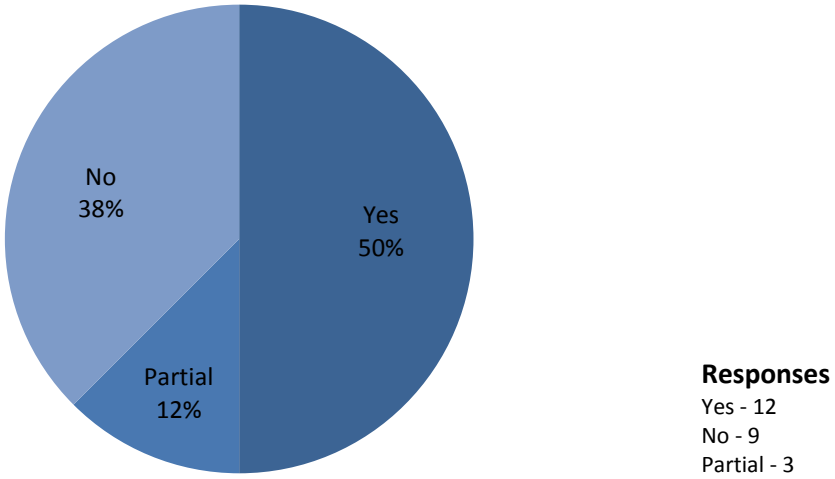
## Section Two: High Value / High Risk Data

Section two covers Q7 and Q8 which sought to identify agency awareness of their high value / high risk data (see Figure E), and the criteria used to assess data as being within these categories (see Figure F).

### Key Findings

The questions around high risk / high value records were addressed by the majority of the respondents. Of the 24 who responded to the question, just over a third (37.5%) indicated that their agency had not identified high value / high risk data, with three (12.5%) reporting some progress.

**Figure E: Identified High Risk / High Value Records**




Where agencies have identified high value / high risk material, a range of criteria, mechanisms and products has been used to create this list. The involvement of information technology and information communication technology in this process is common.

| Figure F: Identification Methods Used |
|---|
| Work Shops |
| Risk assessments |
| ICT Strategy / RM Framework / Quality Control regime development and implementation |
| Business Continuity Strategies and Plan development and implementation |
| System Review |
| Identification by specialist |
| Determining value of the information to the office and government and the impact if the information was lost, unavailable or inadvertently made accessible |
| Disaster Recovery Requirements and User Consultation |
| Assessment against Retention and Disposal Authorities |

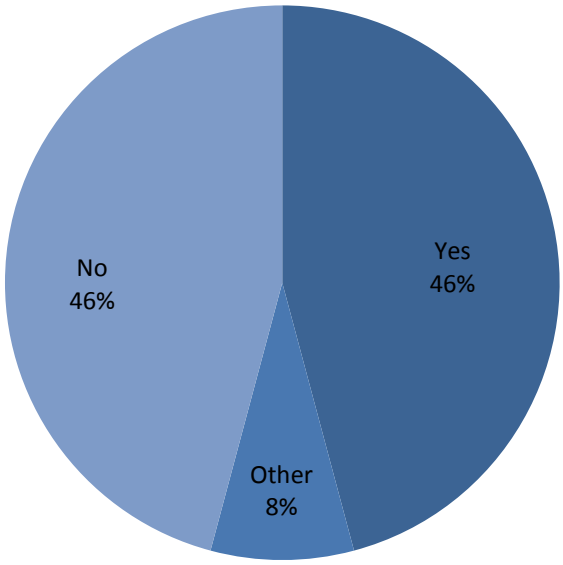
## Section Three: System Redundancy

Section three covers Q9 through to Q12. Q9 asked about systems that were to be or have been decommissioned (see Figure G) and whether they contained data of value. Q10 to Q12 were optional and respondents were only referred to them if they answered Yes to Q9. They were designed to elicit information about known future system redundancy and its impact on records / data.

### Key Findings

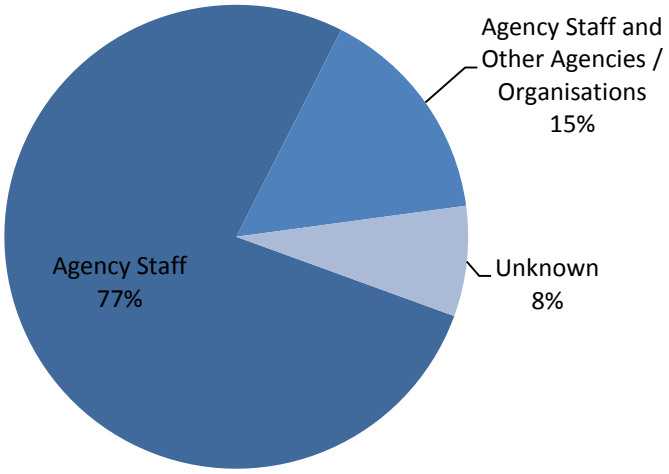
As anticipated, a minority of respondents completed this section, with a majority leaving it blank or answering 'No' to Q9 and hence not seeing the subsequent questions.

**Figure G: Decommissioned Systems**



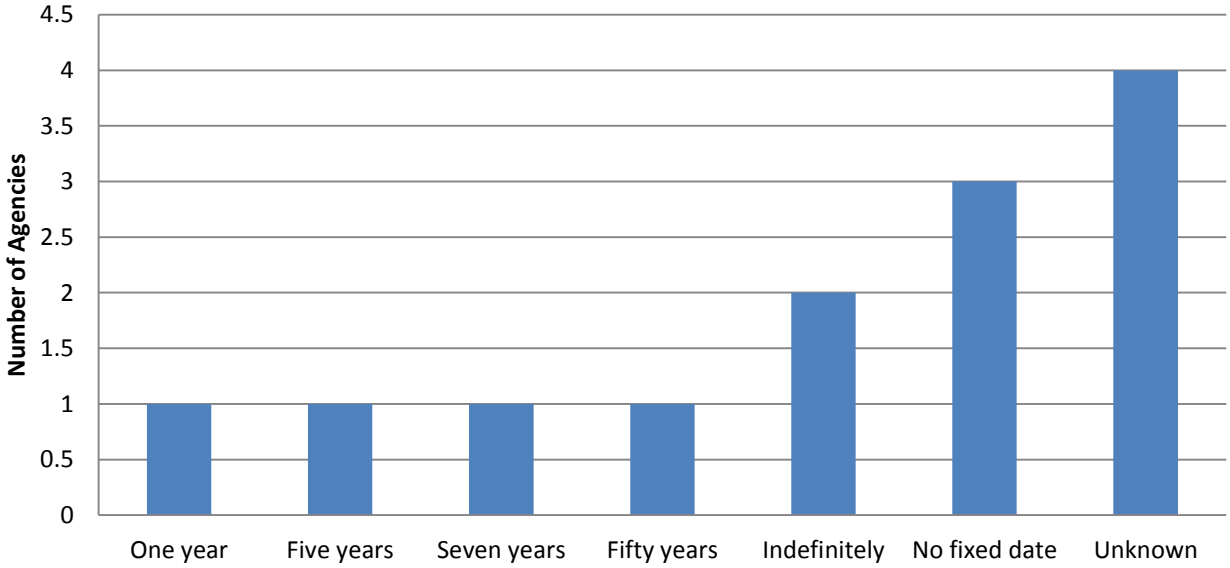

Of the thirteen agencies that responded 'Yes' or 'Other (with further text)' to Q9, the majority indicated that the data from redundant systems will continue to be needed by agency staff, with several also indicating that other agencies will also require ongoing access (see Figure H).

## Figure H: Stakeholders who need the Data




Most agencies do have a plan (or intend to build one) for ensuring ongoing access to the data as part of the shut-down of the old system/s; this typically takes the form of some kind of migration strategy.

## Figure I: Time access is required for




| Figure J: Data Management |
| --- |
| Compliance with Statutory Requirements<br>Migration Plan<br>Active management in system |

## Section Four: Data Storage

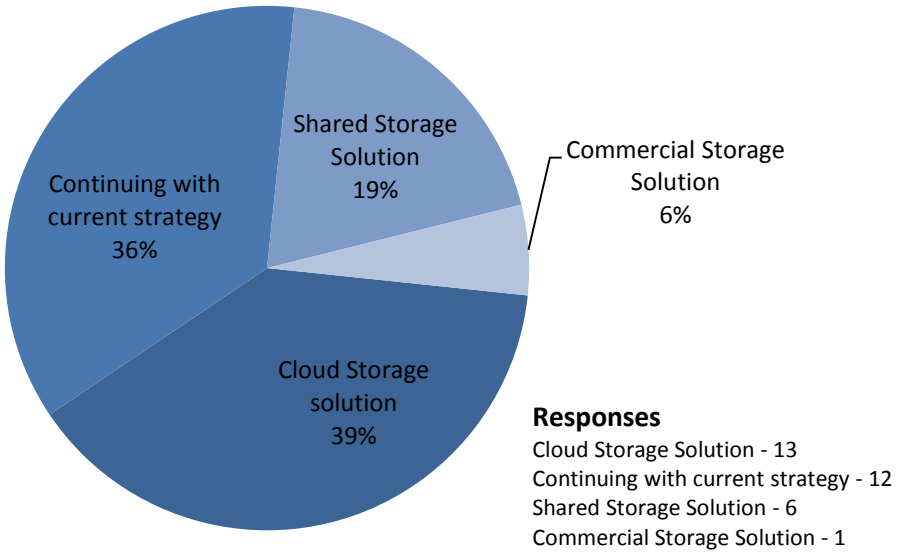
Section 4 covers Q13 regarding data storage and was a drop-down list question where respondents were invited to choose as many options as applied to them (see Figure K).

### Key Findings

Twenty-four respondents made at least one selection in this question. Fifteen respondents (60%) selected only one option or provided explanatory text, while ten (40%) selected more than one option. As the graph (below) shows, the most popular responses were continuation of the current strategy, and moving towards cloud storage solutions.

The question was not sufficiently specific and in-depth enough to ascertain what kind of cloud solutions could be envisaged. This is an area that future surveys could usefully interrogate.

**Figure K: Aggregated Storage Strategies**




Additional comments provided were:

- New strategy is under development with new Chief Information Officer. Most likely a hybrid cloud and local;
- Utilising cloud storage when it is economically viable and our information security requirements can be met; and
- Migration.

## Section Five: Transfer

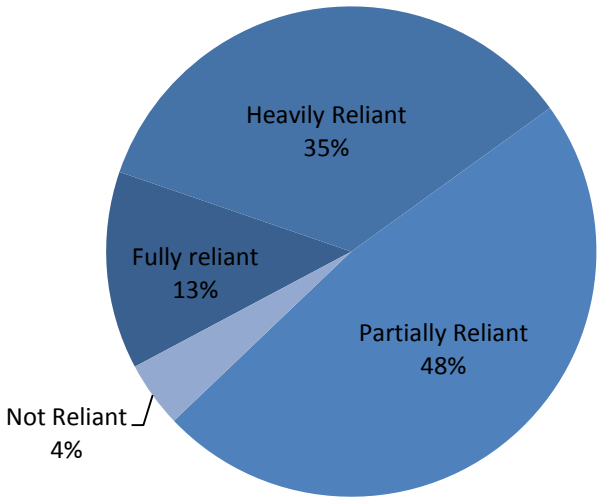
The six questions around retention and disposal and transfer were designed to draw out agencies' understanding of their retention requirements as well as gain a sense of future transfer intentions for digital records.

### Key Findings: Q14 to Q16

The first three questions in this section addressed the continuing use of hardcopy records (see Figure L), coverage / knowledge of RDAs (see Figure M), and knowledge of which electronic records are required for long-term retention (see Figure N). The three questions overlapped and were designed to help clarify whether RDA coverage was applied to all data stores within an agency.

Twenty-three respondents addressed these questions and gave discursive answers to the open field question regarding the level of reliance on hard copy records. Answers were varied, from agencies that indicated very low to no reliance on hard copy through to those which indicated hard copy was still, and was expected to remain, their key recordkeeping format.

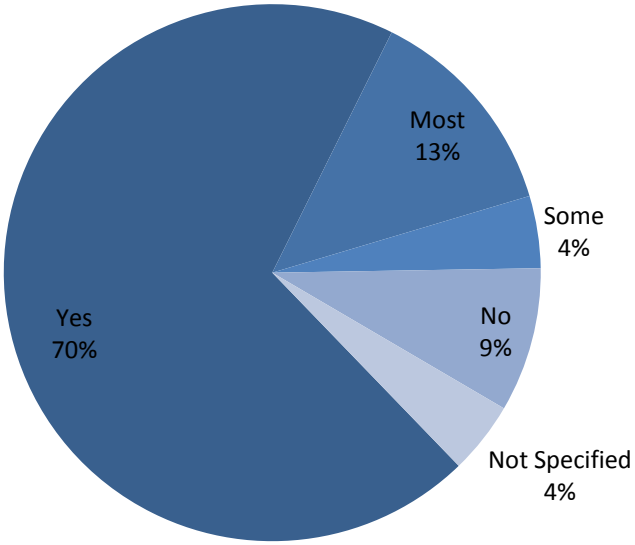
**Figure L: Reliance on hardcopy for business**



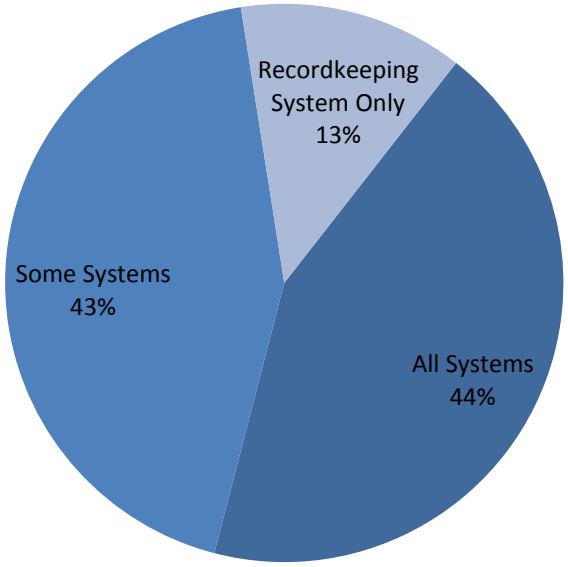

While answers were too diverse to specify one over-riding trend, approximately 50% of respondents indicated that their organisations relied on hard copy only for legacy records or not at all, while approximately 30% indicated total reliance on hard copy, with the remaining 20% indicating that their organisations are in a state of transition towards a planned lower reliance on hard copy in the future.

Two agencies made particular note of statutory requirements that they believe mandated the keeping of hard copy files and restricted their ability to move towards digital models. It is not clear whether this relates to approvals / signatures or some other matter, but it may be a fruitful area for further investigation and advice by PROV.

## Figure M: Covered by a current RDA




## Figure N: Systems with digital records / data required for ten years or longer

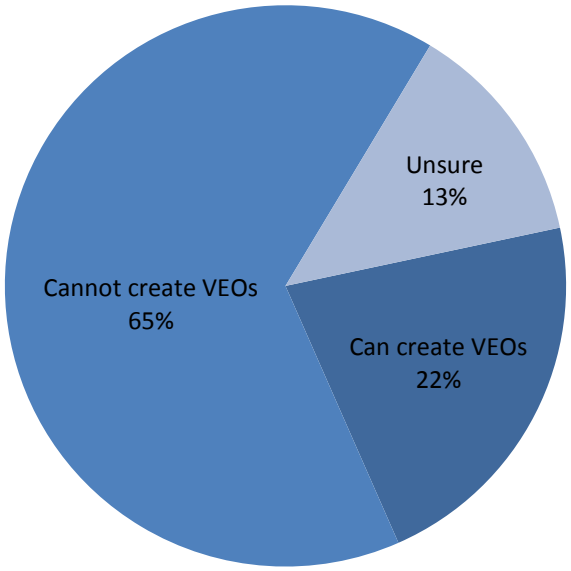



### Key Findings: Q17 to Q19

These questions asked agencies to assess their capacity to generate VEOs (see Figure O) and about the immediate intentions to transfer digital records to PROV (see Figure P). Twenty-three agencies (92%) gave some responses to these questions (see Figure Q).

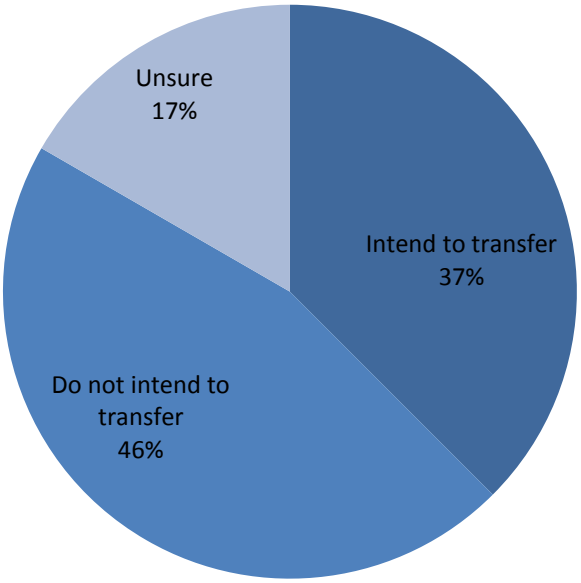
While responses varied, the following representations of the overall responses have been generated. They reveal a possible inconsistency across agencies for PROV, which is that a greater number of agencies have an interest in / intention to transfer digital records than have the capacity to generate VEOs. PROV may need to determine its approach to this issue.

## Figure O: VEO Capacity



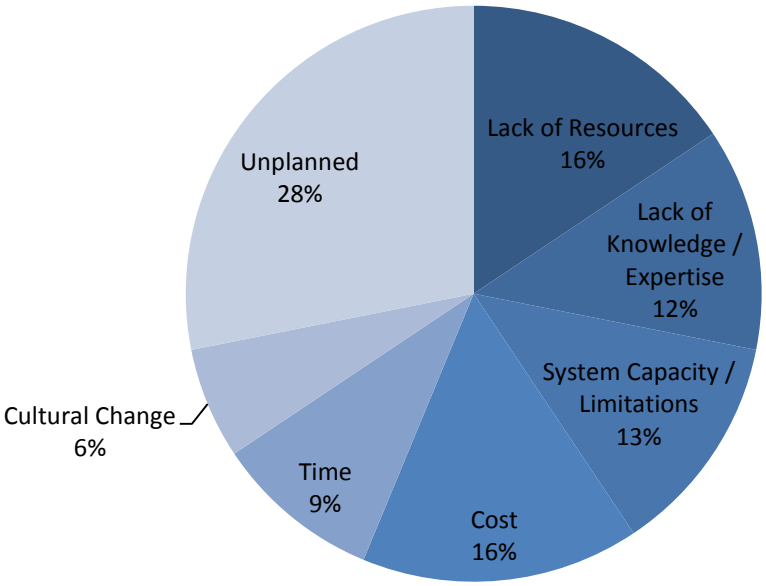

**Responses**
Cannot create VEOS: 15
Can create VEOs: 5
Unsure: 3

## Figure P: Transfer Intentions




**Responses**
Do not intent to transfer: 11
Do intent to transfer: 9
Unsure: 4

## Figure Q: Barriers to Transferring Digital Records




## Section Six: Key Issues

The final question in the survey invited respondents to identify and discuss their top concerns and responses are provided in Figure S (below).

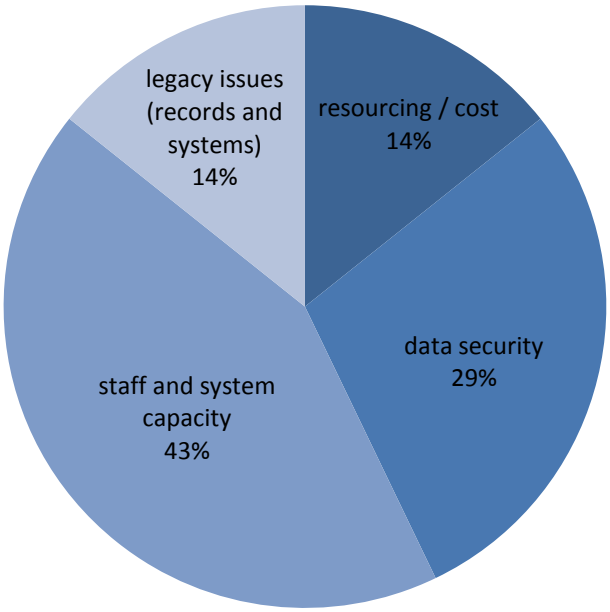
### Key Findings

Nineteen respondents (76%) provided responses to this question. While topics of concern were varied, most repeated were resources / costs, legacy records and systems, staff and system capacity, and data security (see Figure R). Staff and system capacity was the largest response item by a considerable margin, with twelve respondents (48%) indicating concern in this area, followed by data security with eight responses (32%).

Many respondents were willing to, and did, provide detailed information about their challenges and what might be helpful to them in navigating these circumstances.

The results of this question may be useful in providing a base list which PROV could use to poll in greater depth an interested cohort on what areas future advices / policies / standards / training may best be focused.

## Figure R: Top Key Issues




Resourcing / cost: 4
Legacy issues: 4
Data security: 8
Capacity: 12

## Figure S: Key Issues

Awareness of Records and Information Management
Business Continuity
Business Efficiency
Business Systems / Information Stores Management
Capacity
Change Process Implementation / Change management
Cloud Storage / Cloud Computing and software as a service
Compliance not a priority (no motivation / few consequences)
Concept of Information Assets
Consensus building
Consistent application of storage and access controls
Consistently capturing all electronic records with minimum effort by users
Cost
Data governance
Digitisation of Legacy Records / Replacement of hardcopy with electronic
EDRMS / ECM Implementation / EDRMS does not fit need
Ensuring Legislative / Regulatory Framework is Fit for Purpose
Finding a system that is easy for users to store / retrieve data
Increased size of documents
Long Term Preservation
Management of data beyond lifespan of system / Migration of data
Management of disposal of electronic records / Relevant RDA and ease of destruction
Modernising approach to Records Management
Moving from EDRMS to ECM
Navigating records, security and privacy requirements
Need for specialist Records and Information Management Skills / Capacity
Pace of technology and legacy issues with data

| |
|---|
| Reliance on paper<br>Resources<br>Risk of ICT Projects<br>Security / Privacy / Impacts of new Information Security obligations / Security of data<br>Segregation of high and low value electronic records<br>SharePoint Integration<br>Staff Capability and willing participation<br>Strategic Directive / evidence based rationale<br>Storage Capacity<br>Time and resources to implement change / updates<br>Transfer of SharePoint documents to PROV without additional software modifications |