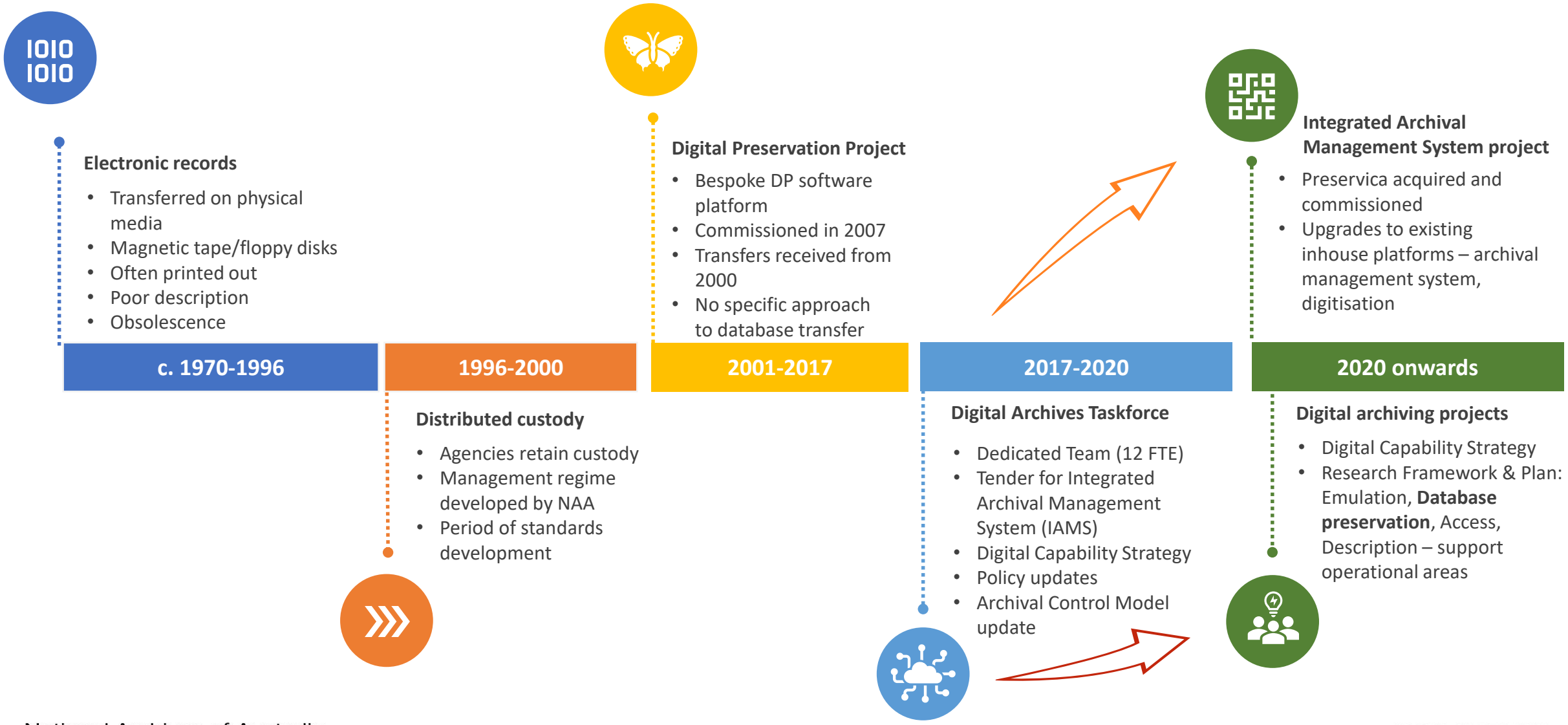
Database Preservation at the National Archives of Australia

NATIONAL ARCHIVES OF AUSTRALIA

What I'll talk about today:

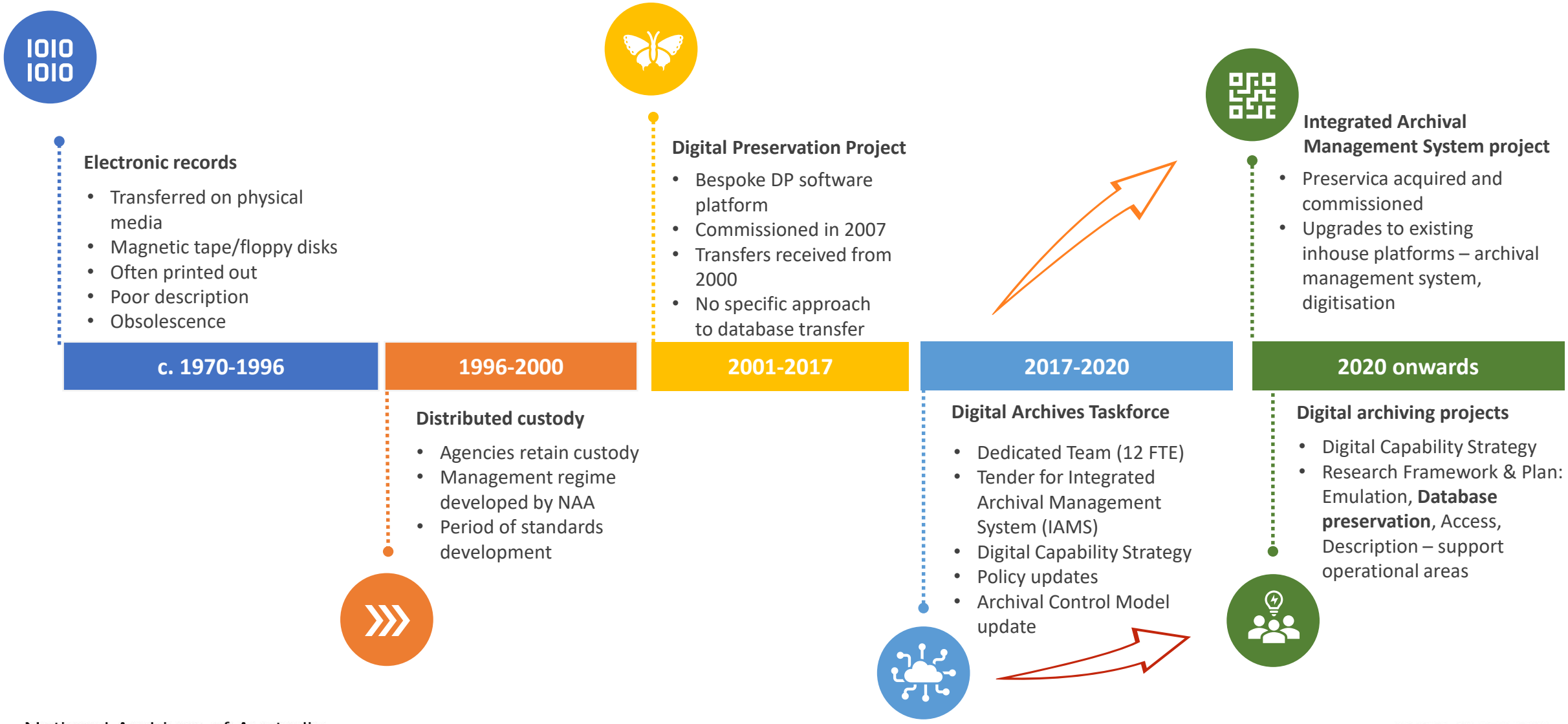
- History of digital records at NAA
- Existing approaches to database transfer and preservation – 2 examples
- Database Preservation Project

A brief history of digital preservation at NAA



AH 1196
AH 1564

A brief history of Digital Preservation at NAA



Current approach at NAA

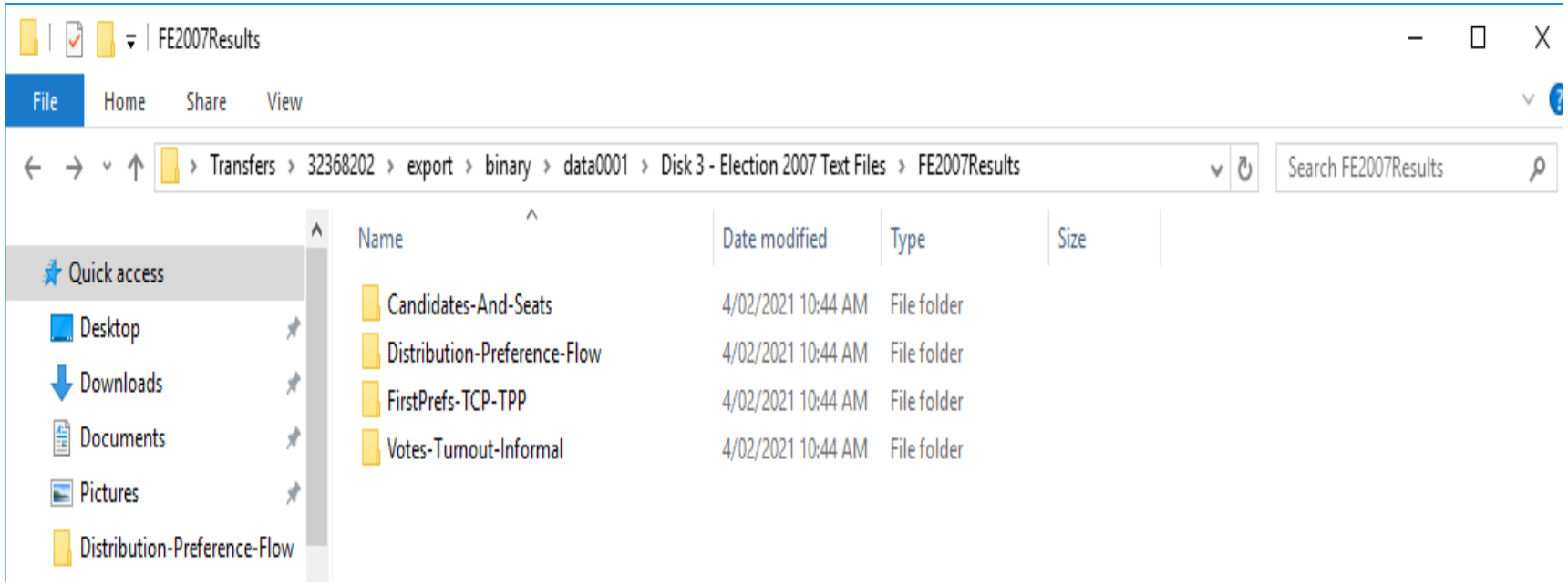
- Not many database transfers received
- Transfers are ad hoc – no standard approach
- We've set a very low bar for agencies (keep it simple)
- Data exports in an open format – plain text, CSV
- Transfer very little technical or structural metadata
- Haven't always asked for data dictionaries or other technical documentation
- Consequently, we have very little information about how these systems were managed

Examples from the NAA collection:
1. 2007 Federal Election Results

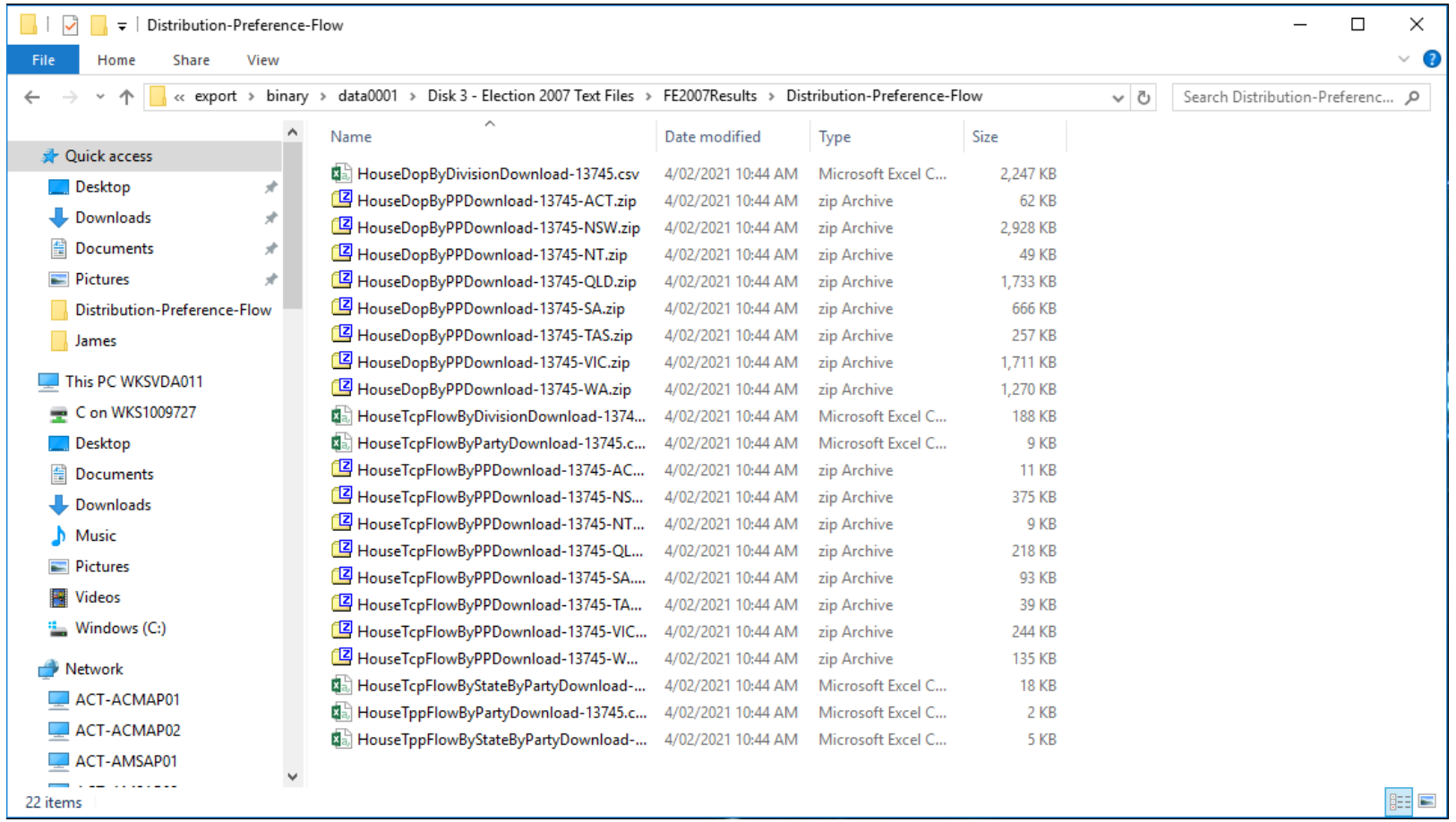
Federal Election Results – Disposal Class for Election Results

Entry	*Description of Records*	*Disposal Action*
Class no. 11729	Final versions of significant reports and returns on the conduct of elections, ballots and referendums. Includes: • official reports on Federal elections presented by the Electoral Commissioner following an election • summary reports on the conduct of referendums • official results of elections or referendum in statistical tables • divisional returning officer returns • election journals.	Retain as national archives

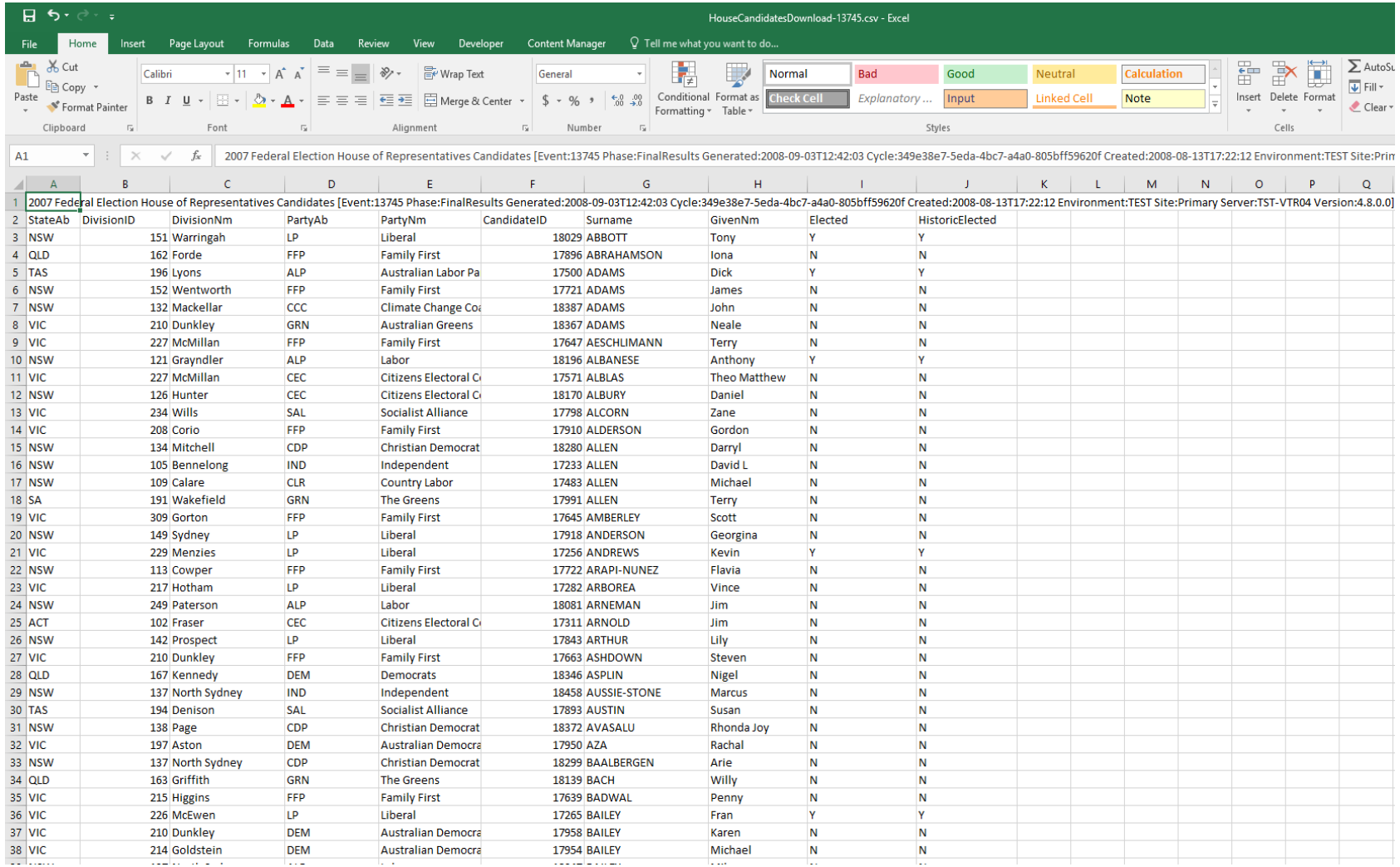
Transferred Data



Transferred Data



Transferred Data



HouseCandidatesDownload-13745.csv - Excel

	A	B	C	D	E	F	G	H	I	J
1	2007 Federal Election House of Representatives Candidates [Event:13745 Phase:FinalResults Generated:2008-09-03T12:42:03 Cycle:349e38e7-5eda-4bc7-a4a0-805bff59620f Created:2008-08-13T17:22:12 Environment:TEST Site:Primary Server:TST-VTR04 Version:4.8.0.0]									
2	StateAb	DivisionID	DivisionNm	PartyAb	PartyNm	CandidateID	Surname	GivenNm	Elected	HistoricElected
3	NSW	151	Warringah	LP	Liberal	18029	ABBOTT	Tony	Y	Y
4	QLD	162	Forde	FFP	Family First	17896	ABRAHAMSON	Iona	N	N
5	TAS	196	Lyons	ALP	Australian Labor Pa	17500	ADAMS	Dick	Y	Y
6	NSW	152	Wentworth	FFP	Family First	17721	ADAMS	James	N	N
7	NSW	132	Mackellar	CCC	Climate Change Coa	18387	ADAMS	John	N	N
8	VIC	210	Dunkley	GRN	Australian Greens	18367	ADAMS	Neale	N	N
9	VIC	227	McMillan	FFP	Family First	17647	AESCHLIMANN	Terry	N	N
10	NSW	121	Grayndler	ALP	Labor	18196	ALBANESE	Anthony	Y	Y
11	VIC	227	McMillan	CEC	Citizens Electoral C	17571	ALBLAS	Theo Matthew	N	N
12	NSW	126	Hunter	CEC	Citizens Electoral C	18170	ALBURY	Daniel	N	N
13	VIC	234	Wills	SAL	Socialist Alliance	17798	ALCORN	Zane	N	N
14	VIC	208	Corio	FFP	Family First	17910	ALDERSON	Gordon	N	N
15	NSW	134	Mitchell	CDP	Christian Democrat	18280	ALLEN	Darryl	N	N
16	NSW	105	Bennelong	IND	Independent	17233	ALLEN	David L	N	N
17	NSW	109	Calare	CLR	Country Labor	17483	ALLEN	Michael	N	N
18	SA	191	Wakefield	GRN	The Greens	17991	ALLEN	Terry	N	N
19	VIC	309	Gorton	FFP	Family First	17645	AMBERLEY	Scott	N	N
20	NSW	149	Sydney	LP	Liberal	17918	ANDERSON	Georgina	N	N
21	VIC	229	Menzies	LP	Liberal	17256	ANDREWS	Kevin	Y	Y
22	NSW	113	Cowper	FFP	Family First	17722	ARAPI-NUNEZ	Flavia	N	N
23	VIC	217	Hotham	LP	Liberal	17282	ARBOREA	Vince	N	N
24	NSW	249	Paterson	ALP	Labor	18081	ARNEMAN	Jim	N	N
25	ACT	102	Fraser	CEC	Citizens Electoral C	17311	ARNOLD	Jim	N	N
26	NSW	142	Prospect	LP	Liberal	17843	ARTHUR	Lily	N	N
27	VIC	210	Dunkley	FFP	Family First	17663	ASHDOWN	Steven	N	N
28	QLD	167	Kennedy	DEM	Democrats	18346	ASPLIN	Nigel	N	N
29	NSW	137	North Sydney	IND	Independent	18458	AUSSIE-STONE	Marcus	N	N
30	TAS	194	Denison	SAL	Socialist Alliance	17893	AUSTIN	Susan	N	N
31	NSW	138	Page	CDP	Christian Democrat	18372	AVASALU	Rhonda Joy	N	N
32	VIC	197	Aston	DEM	Australian Democra	17950	AZA	Rachal	N	N
33	NSW	137	North Sydney	CDP	Christian Democrat	18299	BAALBERGEN	Arie	N	N
34	QLD	163	Griffith	GRN	The Greens	18139	BACH	Willy	N	N
35	VIC	215	Higgins	FFP	Family First	17639	BADWAL	Penny	N	N
36	VIC	226	McEwen	LP	Liberal	17265	BAILEY	Fran	Y	Y
37	VIC	210	Dunkley	DEM	Australian Democra	17958	BAILEY	Karen	N	N
38	VIC	214	Goldstein	DEM	Australian Democra	17954	BAILEY	Michael	N	N

Series Registration for A13777: 2007 Federal Election Results

Series details for: A13777	
Series number	A13777
Title	Federal Election results
Accumulation dates	1901 -
Contents dates	2007 -
Items in this series on RecordSearch	4 All items from this series are entered on RecordSearch.
Agency/person recording	21 Feb 1984 - CA 3854, Australian Electoral Commission, Central Office
Agency/person controlling	CA 3854, Australian Electoral Commission, Central Office
Quantity and location	0.2 metres held in ACT
System of arrangement/ control	Single number series
Predominant physical format	ELECTRONIC RECORDS
Series note	**Summary heading** **Function and purpose** **Related legislation** **Using the series** **Language of material** **Physical characteristics** **System of arrangement and control** **Relationships with other records** **Finding aids** **Access conditions** **Series history** **Provenance** **Immediate source of acquisition** **Custodial history** **Quantity in agency custody** **Disposal history** **Publication note** **Additional information** **End notes** **Sources** (Less)
Visibility & availability indicator	73 . All items from the series are on RecordSearch
Date registered	29 Jun 2009
Date modified	24 Jul 2014

Item Registration for A13777

Item details for: A13777, 3	Issue to research centre Request copy
Title	Federal Election Results - 2007 - text files
Contents date range	2007 - 2007
Series number	A13777 Click to see which government agency or person created this item.
Control symbol	3
Citation	NAA: A13777, 3
Item ID	32368202
Location	Canberra
Access status	Not yet examined
Physical format	ELECTRONIC RECORDS (allocated at series level)
Series accession number	2009/00555361
Records authority class number	11729
Date registered	24 Jul 2014
Date modified	24 Jul 2014

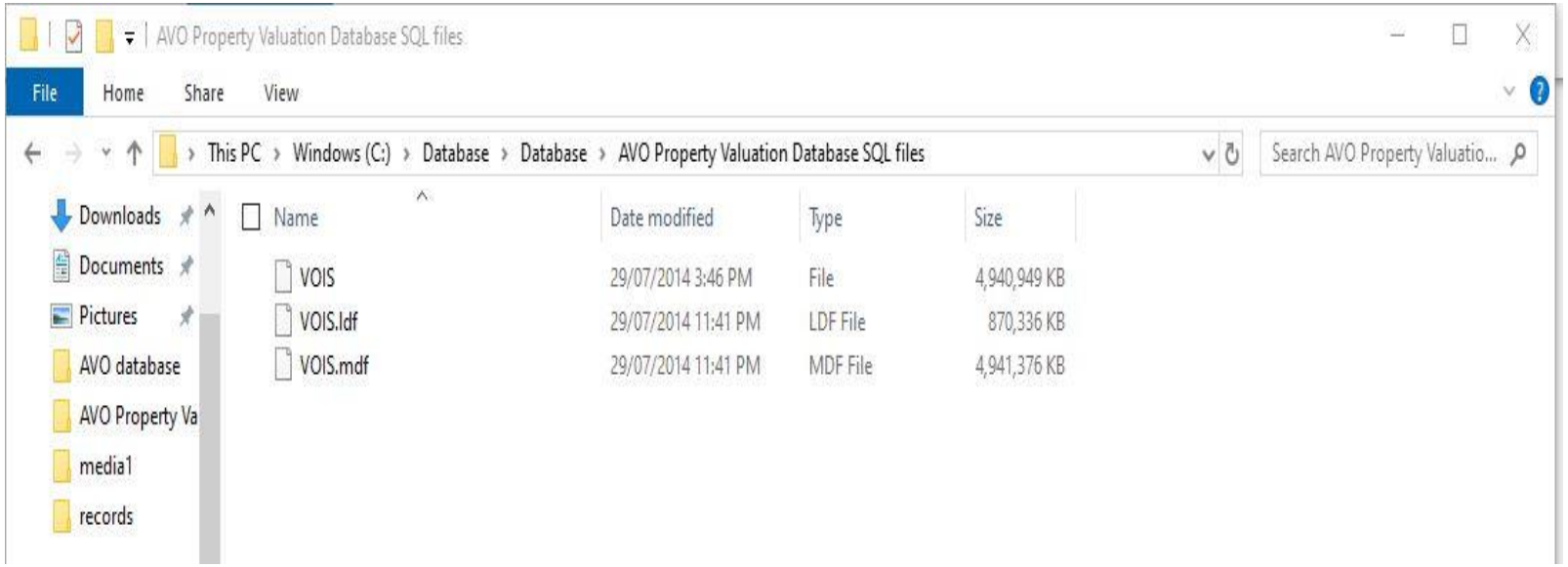
Issues

- Description
- Technical documentation
- Raw data format
- You can probably think of other issues

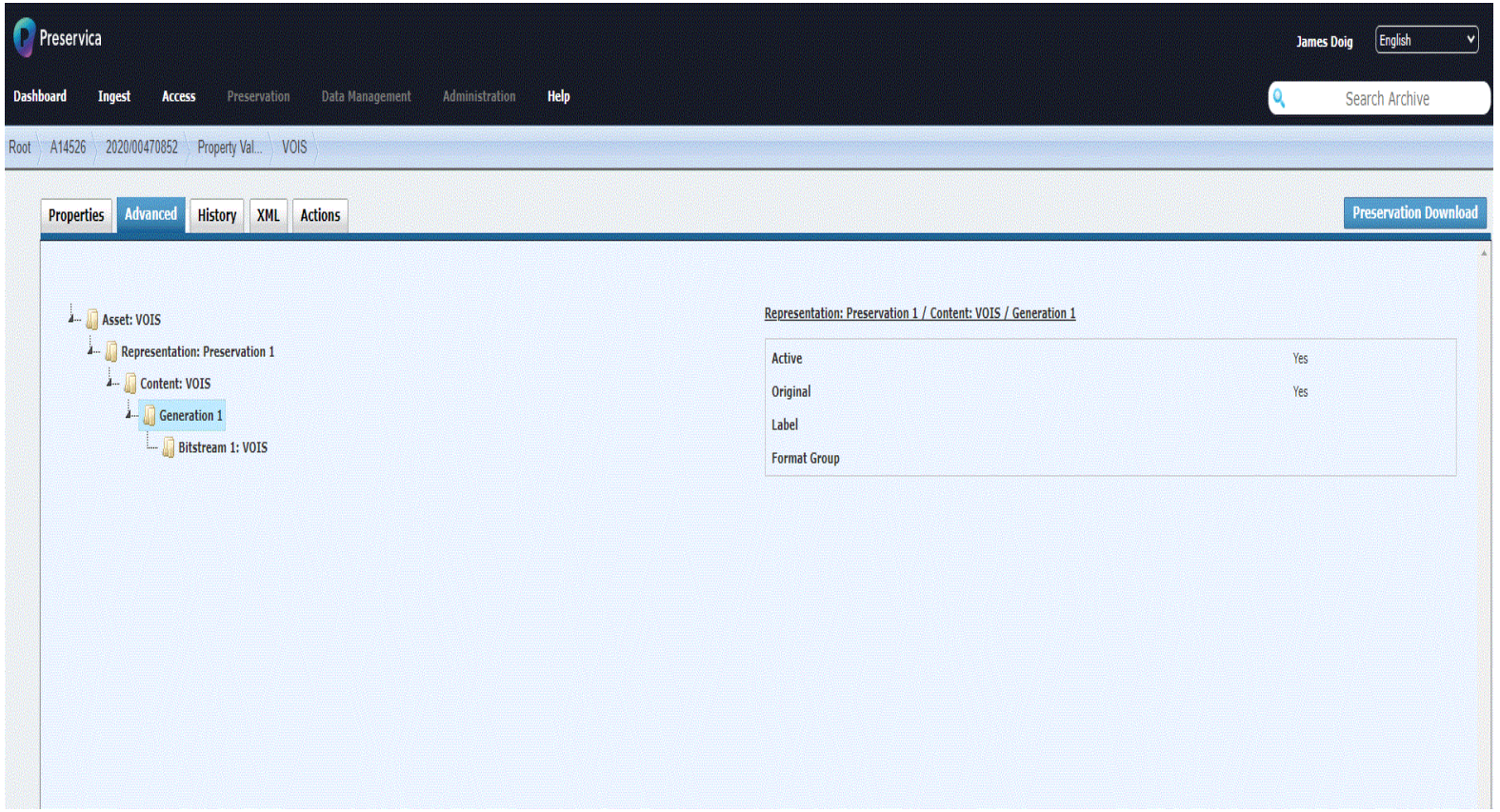
Examples from the NAA collection:

2. Australian Valuation Office Database (2014)

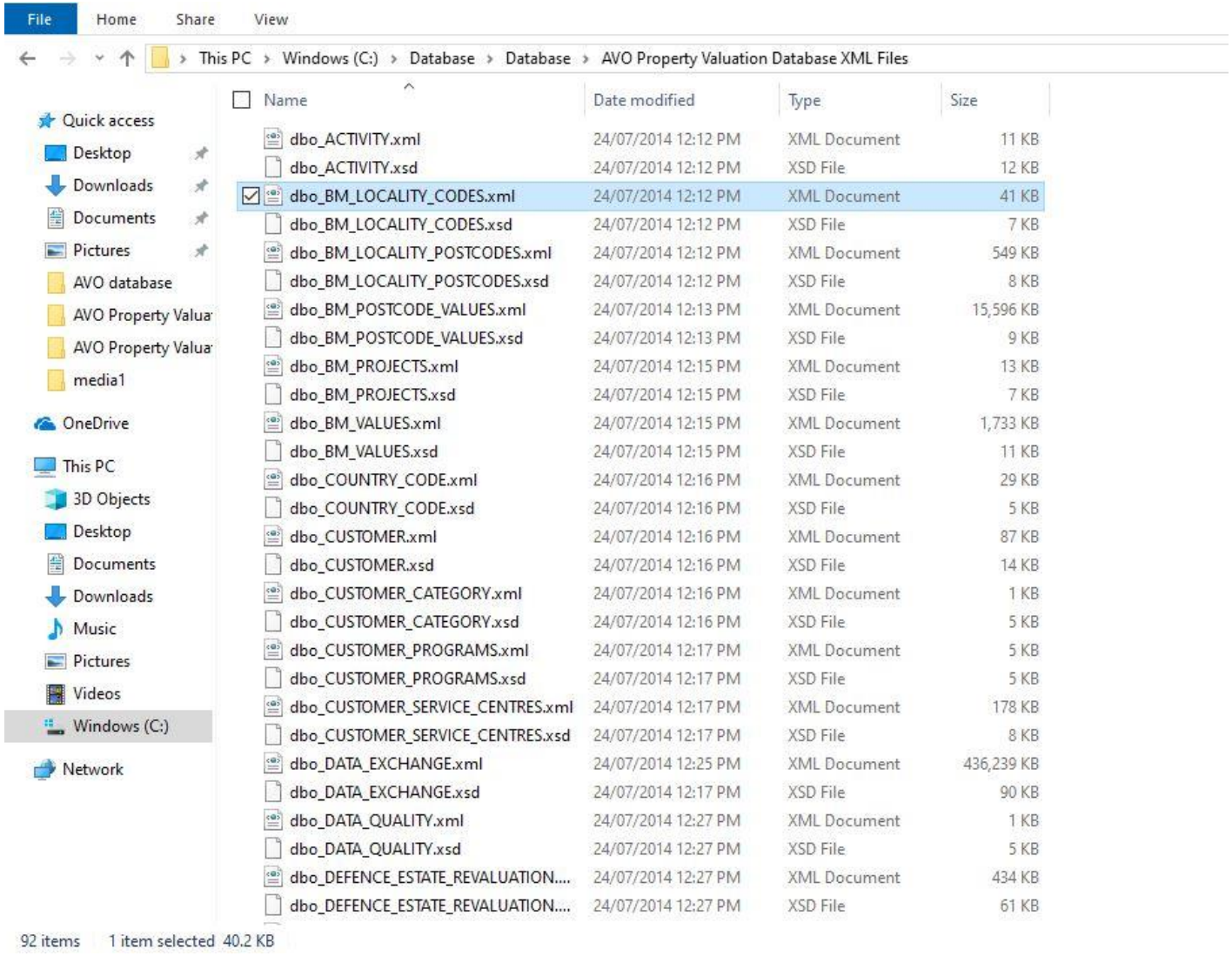
Native SQL files



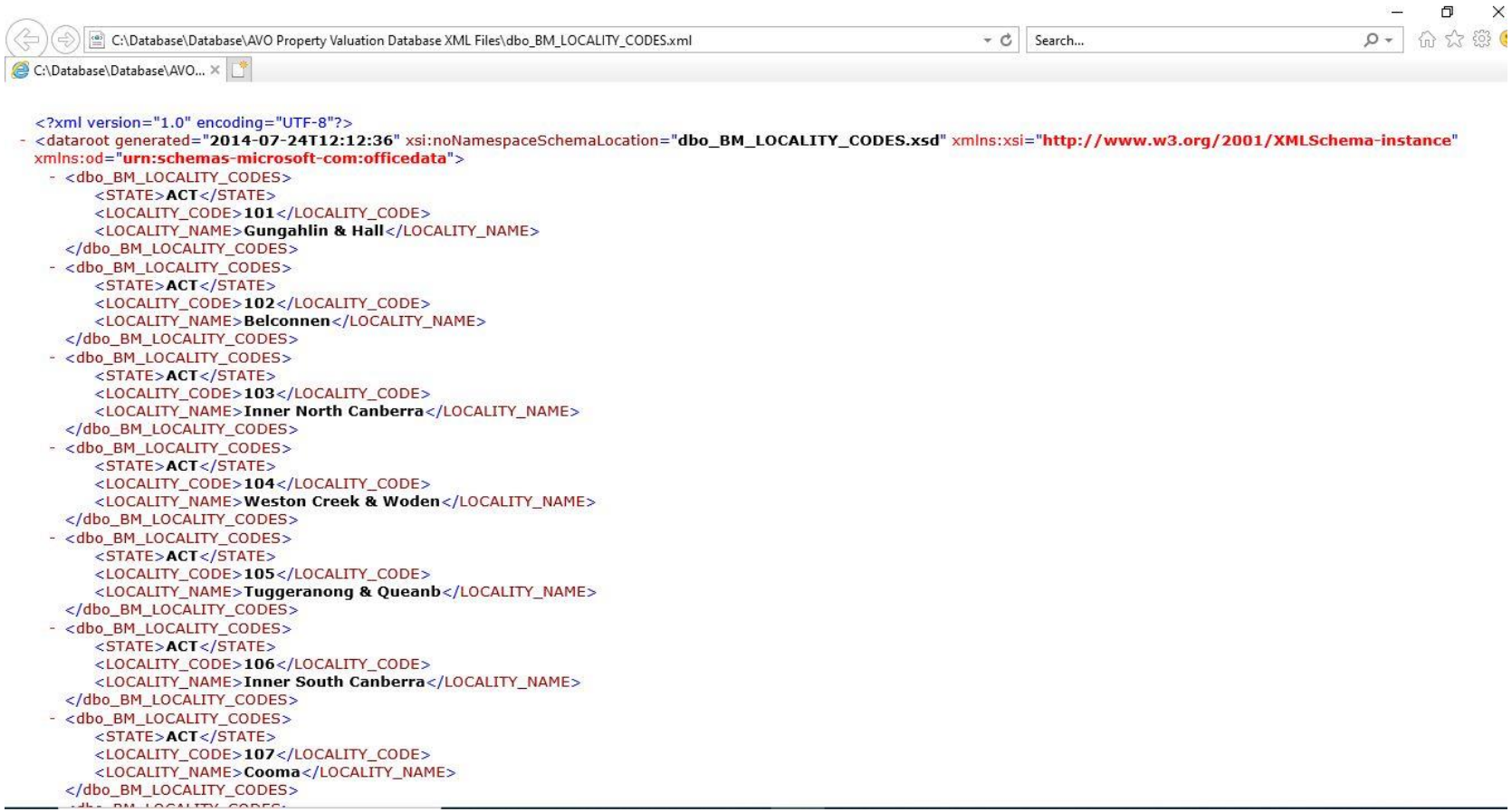
Preservica file format identification



XML files of database tables



XML file of a database table




```
<?xml version="1.0" encoding="UTF-8"?>
- <dataroot generated="2014-07-24T12:12:36" xsi:noNamespaceSchemaLocation="dbo_BM_LOCALITY_CODES.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:od="urn:schemas-microsoft-com:officedata">
  - <dbo_BM_LOCALITY_CODES>
      <STATE>ACT</STATE>
      <LOCALITY_CODE>101</LOCALITY_CODE>
      <LOCALITY_NAME>Gungahlin & Hall</LOCALITY_NAME>
    </dbo_BM_LOCALITY_CODES>
  - <dbo_BM_LOCALITY_CODES>
      <STATE>ACT</STATE>
      <LOCALITY_CODE>102</LOCALITY_CODE>
      <LOCALITY_NAME>Belconnen</LOCALITY_NAME>
    </dbo_BM_LOCALITY_CODES>
  - <dbo_BM_LOCALITY_CODES>
      <STATE>ACT</STATE>
      <LOCALITY_CODE>103</LOCALITY_CODE>
      <LOCALITY_NAME>Inner North Canberra</LOCALITY_NAME>
    </dbo_BM_LOCALITY_CODES>
  - <dbo_BM_LOCALITY_CODES>
      <STATE>ACT</STATE>
      <LOCALITY_CODE>104</LOCALITY_CODE>
      <LOCALITY_NAME>Weston Creek & Woden</LOCALITY_NAME>
    </dbo_BM_LOCALITY_CODES>
  - <dbo_BM_LOCALITY_CODES>
      <STATE>ACT</STATE>
      <LOCALITY_CODE>105</LOCALITY_CODE>
      <LOCALITY_NAME>Tuggeranong & Queanb</LOCALITY_NAME>
    </dbo_BM_LOCALITY_CODES>
  - <dbo_BM_LOCALITY_CODES>
      <STATE>ACT</STATE>
      <LOCALITY_CODE>106</LOCALITY_CODE>
      <LOCALITY_NAME>Inner South Canberra</LOCALITY_NAME>
    </dbo_BM_LOCALITY_CODES>
  - <dbo_BM_LOCALITY_CODES>
      <STATE>ACT</STATE>
      <LOCALITY_CODE>107</LOCALITY_CODE>
      <LOCALITY_NAME>Cooma</LOCALITY_NAME>
    </dbo_BM_LOCALITY_CODES>
```

AVO Data Dictionary

Australian Valuation Office Data Dictionary

Contents

Australian Valuation Office Data Dictionary 4
Introduction 7
Audience 7
Purpose 7
Contact Information 8

The AVODD 9
Introduction 9
Logical structure of Appendable Tables 10
Points related to the AVODD 10
Bench Mark Tables 11
Sample SQL and Joins for Main Tables 12
Locating foreign keys from the filterable excel version of the tabulations. 12
The JOBS table and related simple look up table joins 13
The PROPERTY_ASSET_VAL table and related simple look up table joins 15
The BENCH MARK TABLES and related joins 16
TIME_SPENT table and related joins 17

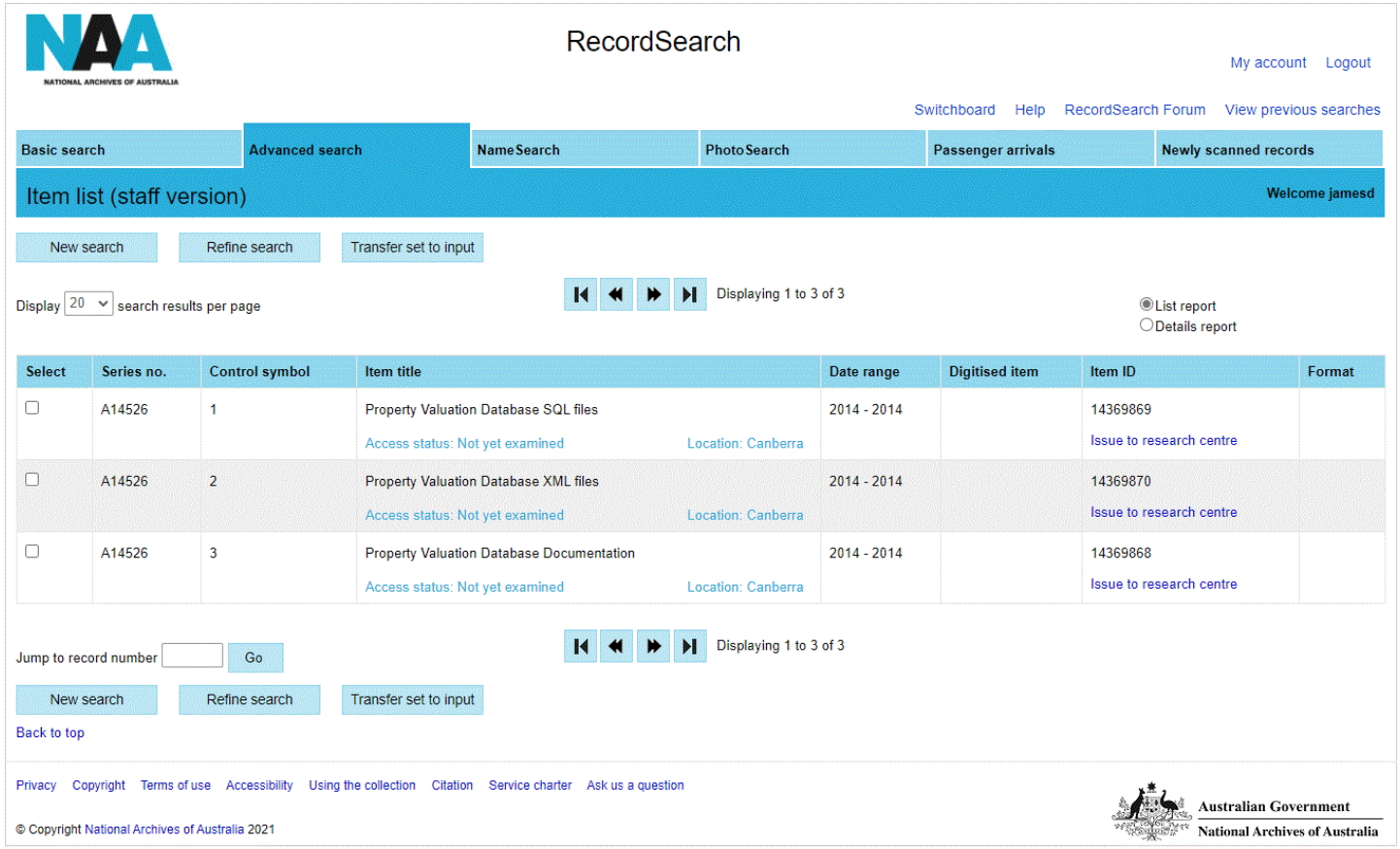
The Data Tables and Data Elements 18
ACCESS_LEVEL 18
ACTIVITY 18
AREA_POSTCODES 19
ATO_AVO_GATEKEEPERS 20
AVO_ADDRESS 20
AVO_GROUP 22
AVO_SKILLS 22
BM_LOCALITY_CODES 23
BM_LOCALITY_POSTCODES 23
BM_LOCALITY_POSTCODES_OLD 24
BM_POSTCODE_VALUES 25
BM_PROJECTS 27
BM_PROPERTY_TYPES 27
BM_VALUES 28
BOOKING_CATEGORY 29
BULK_JOB_CODE 29
BULK_LODGEMENT_DEFAULTS 30
CLK_DATA 31
CLK_FEES 35
CONFERENCE_BOOKING_DETAILS 35
CONFERENCE_ROOMS 36
COUNTRY_CODE 36
CUST_REGION 37
CUSTOMER 37
CUSTOMER_CATEGORY 39
CUSTOMER_PROGRAMS 39

CUSTOMER_SERVICE_CENTRES 40
DarwinViewAllJobs 41
DarwinViewOnHand 43
DarwinViewOnHandPE_Other 45
DarwinViewWithValues 49
DATA_EXCHANGE 53
DATA_EXCHANGE_DATE_IMPORTED 60
DATA_QUALITY 60
DataExchangeDisp10Oct13 61
DEFAULT_TIME 68
DEFENCE_ESTATE_REVALUATION 69
DEFENCE_ESTATE_REVALUATION_OLD 74
DEFENCE_PROJECT_STAGE 80
DEFENCE_REJECT_INFO 80
DEFENCE_REJECT_INFO_OLD 81
DTPROPERTIES 81
DVA_DATA 82
DVA_FEES 85
EVAL_SALES 86
FAT_LODGEMENT_DEFAULTS 87
HOURS_WORKED 88
HOUSE_CURTILAGE_TYPE 90
INVOICEFEESSNAPSHOT 90
JOBS 92
JOBS_ASSET_TITLE 95
JOBS_ASSET_VAL 98
JOBS_ASSET_VAL_HISTORICS 103
JOBS_OTHERS_VAL 104
JOBS_PE_VAL 105
JOBSCOPY 107
LEAVE_APPL 111
LIAISON_INDICATOR 113
LOCATION 113
LOGIN_LOG 114
LOOKUP_TABLES 114
MONTHLY_REPORT 115
MOU_TIMEFRAMES 127
NEW_AREA_POSTCODES 128
NOTICE_BOARD 128
OFFICE 128
OFFICER 129
OFFICER_GRADE 133
OFFICER_GROUP 134
OFFICER_LEVEL 134
OFFICERBAK 135
PRIORITY 139
PRODUCT_LINE 139
PROJECT 139
PROJECT_CUSTOMER_CODE 140
PROPERTY 141
PROPERTY_ASSET_VAL 143
PROPERTY_CONDITION 144
PROPERTY_EXTERIOR 144
PROPERTY_INTERIOR 145
PROPERTY_ROOF 145

Series Registration for A14526

Series details for: A14526	
Series number	A14526
Title	Australian Valuation Office Database, 2014
Accumulation dates	14 Jan 2014 - 29 Jul 2014
Contents dates	14 Jan 2014 - 29 Jul 2014
Items in this series on RecordSearch	3 Click to see items listed on RecordSearch. Please contact the National Reference Service if you can't find the record you want as not all items from the series may be on RecordSearch.
Agency/person recording	14 Jan 2014 - 30 Jun 2014 CA 7089, Australian Valuation Office, Central Office 01 Jul 2014 - 29 Jul 2014 CA 995, Australian Taxation Office, Head Office
Agency/person controlling	01 Jul 2014 - CA 995, Australian Taxation Office, Head Office
Quantity and location	0.1 metres held in ACT
System of arrangement/ control	Single number imposed by the National Archives
Range of control symbols	[1] to [3]
Predominant physical format	ELECTRONIC RECORDS
Series note	**Function and purpose** This Series is the legacy valuation database of the Australian Valuation Office called VOIS (Valuation Office Information System). The Series comprises three items. One item consists of technical database documentation including the data dictionary, which explains the structure of the database and the meaning of the data it contains. One item contains the raw SQL data in .mdf (SQL Server data file) and .ldf (SQL Server log file) format. The third item contains the data exported to XML format. The VOIS database recorded all valuation data collected by the AVO. The VOIS database has 106 tables, several of which were in use in the decade up to 2014, when the database was decommissioned. Some of the VOIS tables were out of date or obsolete as of 2014, but were still retained in the database and are of historical interest. There are also some tables that were in sporadic use up to the year 2000 or earlier. The 106 tables comprise over 13,000,000 records. **Related legislation** **Using the series** **Language of material** **Physical characteristics** **System of arrangement and control** Single running number imposed by the National Archives **Relationships with other records** **Finding aids** **Access conditions** **Series history** **Provenance** **Immediate source of acquisition** These electronic records were sent to the National Archives when the Australian Valuation Office (AVO) was closed on 30 June 2014. **Custodial history** **Quantity in agency custody** **Disposal history** **Publication note** **Additional information** **End notes** **Sources** (Less)
Date registered	20 Nov 2020
Date modified	20 Nov 2020

Items in A14526



NAA
NATIONAL ARCHIVES OF AUSTRALIA

RecordSearch

My account Logout

Switchboard Help RecordSearch Forum View previous searches

Basic search | Advanced search | NameSearch | PhotoSearch | Passenger arrivals | Newly scanned records

Item list (staff version) Welcome jamesd

New search Refine search Transfer set to input

Display 20 search results per page

Displaying 1 to 3 of 3

List report
Details report

Select	Series no.	Control symbol	Item title	Date range	Digitised item	Item ID	Format
☐	A14526	1	Property Valuation Database SQL files Access status: Not yet examined Location: Canberra	2014 - 2014		14369869 Issue to research centre	
☐	A14526	2	Property Valuation Database XML files Access status: Not yet examined Location: Canberra	2014 - 2014		14369870 Issue to research centre	
☐	A14526	3	Property Valuation Database Documentation Access status: Not yet examined Location: Canberra	2014 - 2014		14369868 Issue to research centre	

Jump to record number Go

Displaying 1 to 3 of 3

New search Refine search Transfer set to input

Back to top

Privacy Copyright Terms of use Accessibility Using the collection Citation Service charter Ask us a question

© Copyright National Archives of Australia 2021

Australian Government
National Archives of Australia

Problem is…

- Disconnect between development of disposal schedule and transfer
- We haven't decided what we need to preserve to ensure meaningful access in the future.
- The raw data may not be enough:
  - Native SQL format is software dependent
  - Flat file format (e.g. CSV) has limited usability
  - Need technical documentation to understand data
- Are there any characteristics of the database that we need to preserve to ensure meaningful accessibility and usability?
- There is not be a one-size-fits-all approach to database transfer and preservation.

Database Preservation Project: December 2020-July 2021

“to develop and agreed approach to the appraisal, transfer, preservation and management of database, and to develop guidance advice and other supporting artefacts to support that approach.”

Product 1: SIARD Case Study

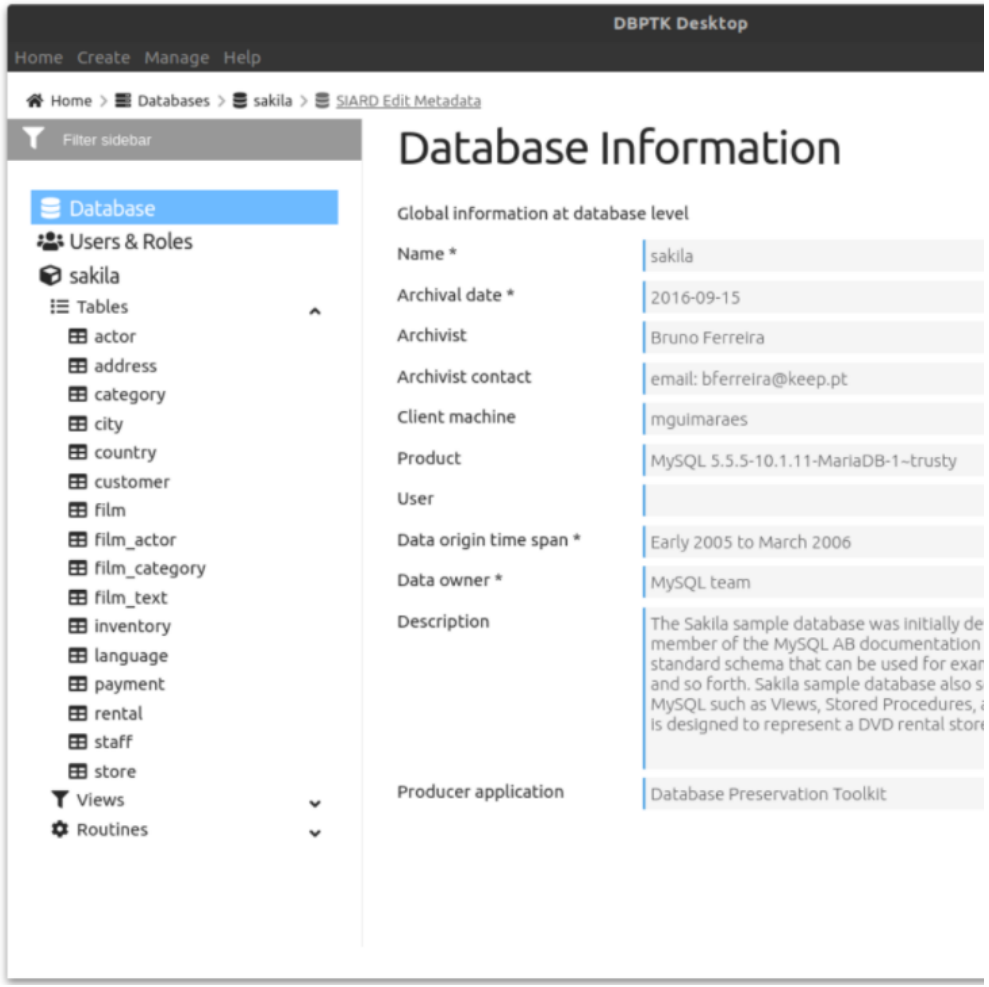
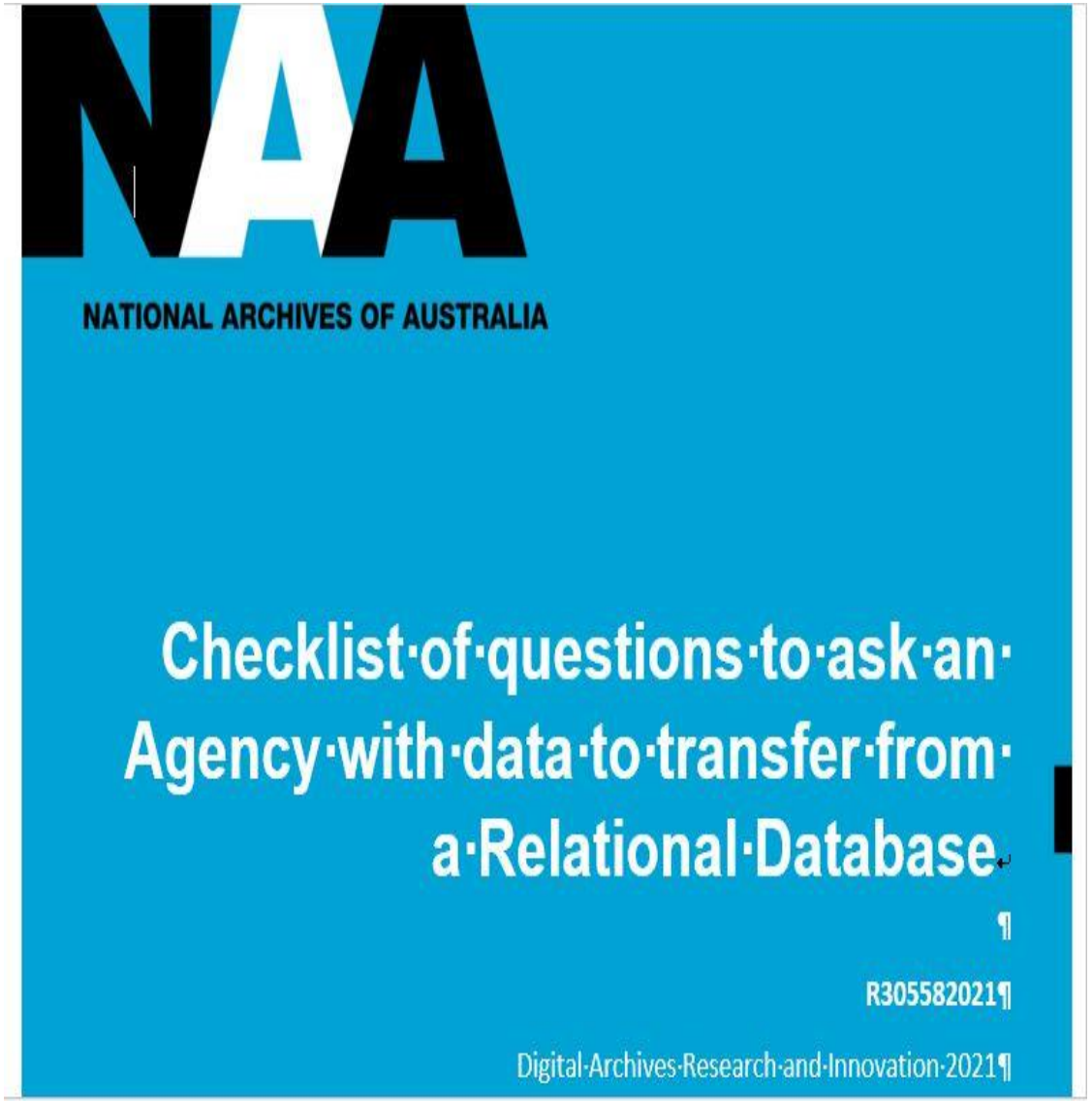


Image Source: https://database-preservation.com/#desktop

Valuation Office Information System

- 2014 – Closed agency transfer
  - Data file in native SQL (.MDF)
  - Log File in native SQL (.LDF)
  - Table data individually exported in XML
  - Technical documentation (Data Dictionary)
  - Screenshots of UI
- 2021 – SIARD file



4. Using this checklist

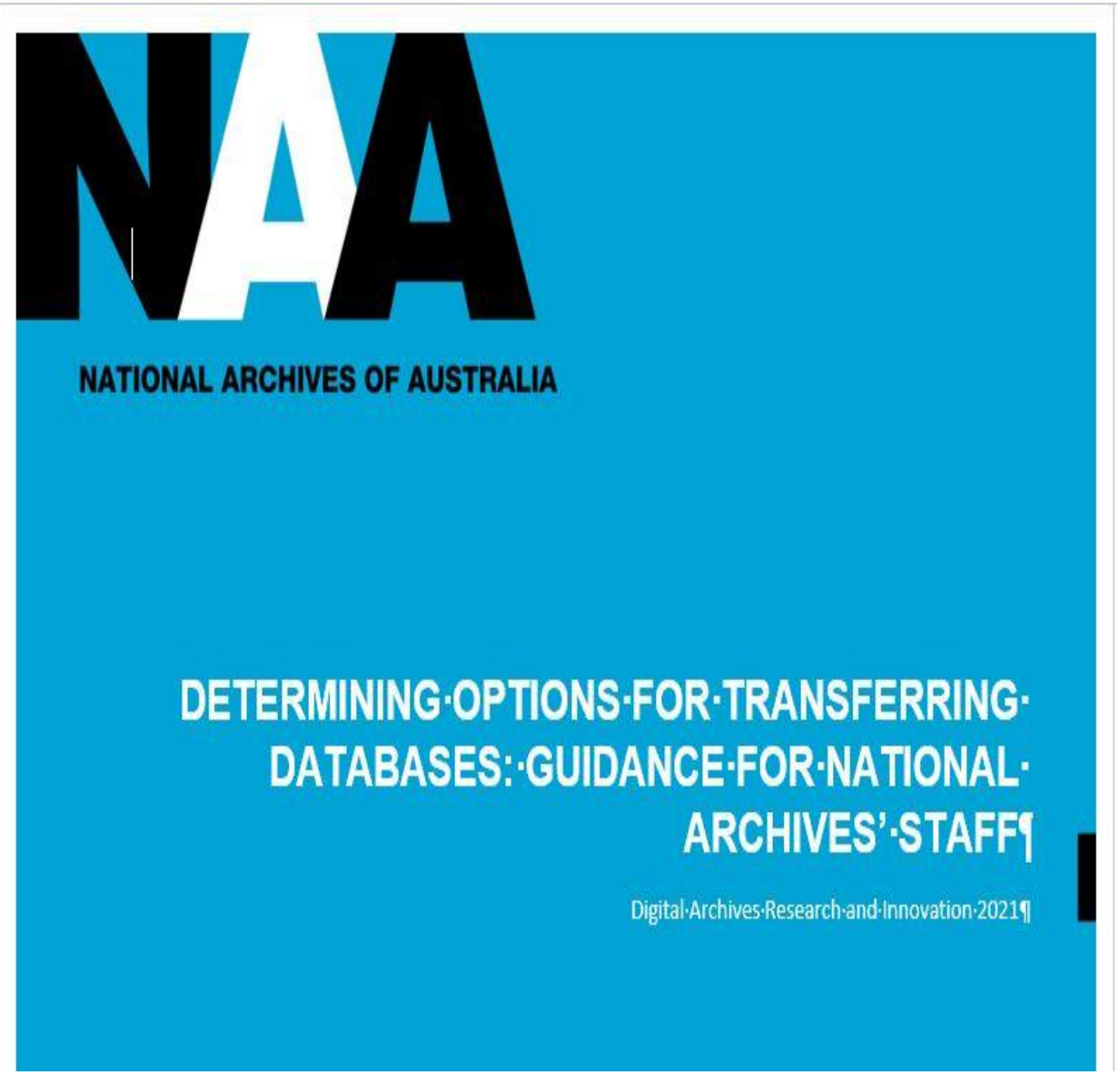
The checklist is intended to assist National Archives' staff to guide discussions with agencies to come to a shared understanding of the selection and transfer process, i.e. appraisal, pre-ingest and ingest and will not be comprehensive for every system or every transfer.

While the checklist is detailed, it is important to note there are four main questions that will shape the form the transfer takes and which will influence which questions to focus the conversation on, and which will shape the transfer.

1. Is the entire database RNA or only part of it?
2. If only part is RNA, does the system have the capacity to export that content?
3. If a partial export is not possible or it is all RNA, can it export the entire dataset?
4. Establish if periodic snapshots are required, i.e. does the system retain all changes and have roll-back capability, or is the content subject to ongoing change? If subject to change, eg data is overwritten or permanently deleted, periodic transfer will be necessary.

The investigation may involve several meetings involving different staff and roles, especially for more complex databases/business systems. Initial meetings/contact usually take place with the agency IM manager/contact. Later meetings will include IT and will discuss system functionality/capability in detail.

It is recommended that the information gathered in response to these questions is collaboratively documented and shared between National Archives and the agency, so it can be built on as the transfer progresses.



Determining·frequency·of·transfer¶

This·section·lists·criteria·for·determining·whether·a·transfer·of·data·contained·in·a·database·is·a·single·one-off·transfer·or·a·periodic·snapshot·carried·out·on·an·agreed·frequency.¶

One-off·transfer¶

One-off·transfers·(or·end-of-life·transfers)·are·usually·made·in·the·following·situations:¶

- Closed·agency¶
- Machinery·of·Government·change,·for·example·the·inheriting·agency·may·decide·to·transfer·the·database·to·the·National·Archives¶
- Decommissioning·of·system¶
- At·the·point·of·migration·to·a·new·system·or·platform¶
- Expiry·of·software·or·system·licence·or·support¶

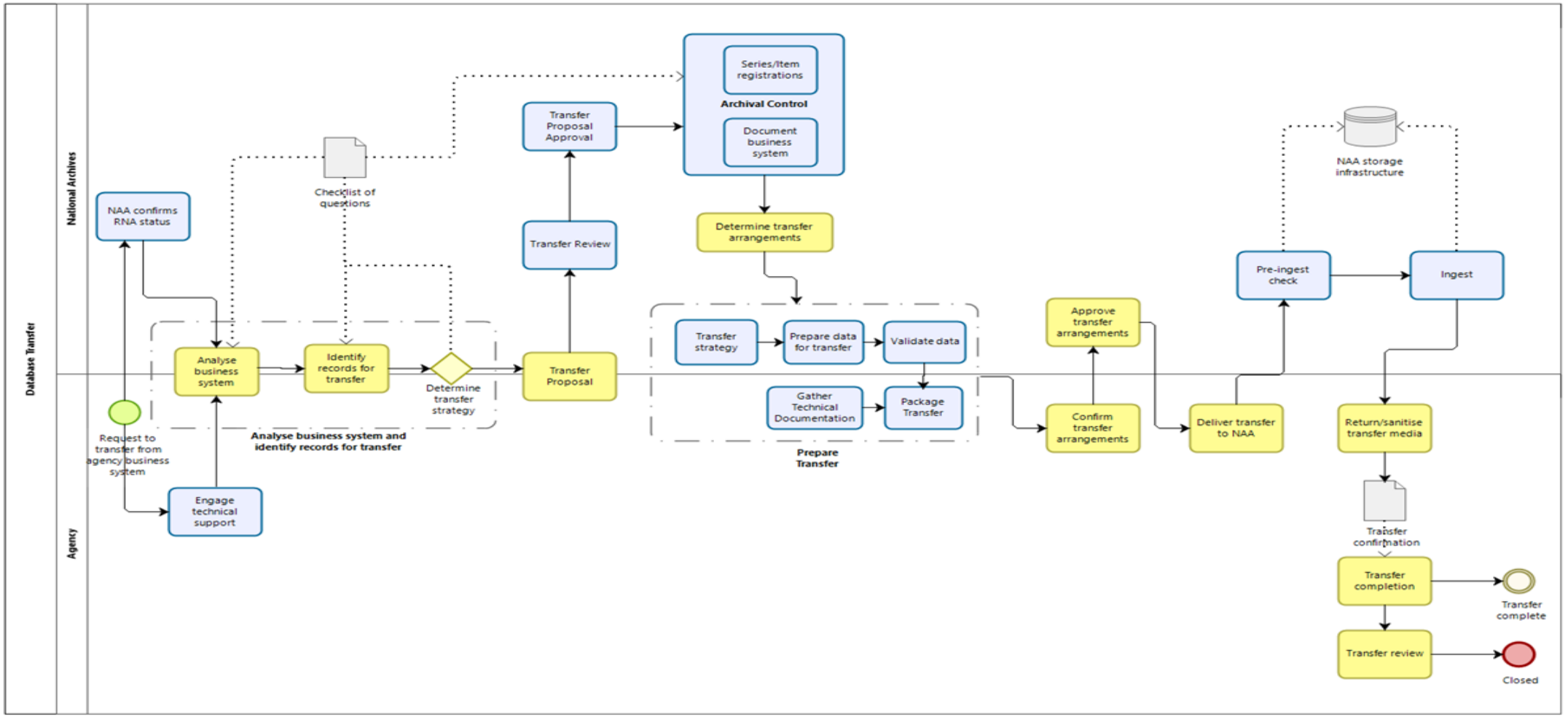
Periodic·snapshot¶

Periodic·snapshots·are·made·in·the·following·situations·and·need·to·be·appropriately·planned·and·managed:¶

- Business·system·periodically·deletes·or·overwrites·data·without·capturing·an·accurate·and·complete·log·of·data·changes¶
- Business·system·is·highly·dynamic¶
- Business·system·has·on-going·business·use,·but·high·value·or·risk·has·led·to·a·decision·to·transfer·periodically·to·the·national·Archives.¶

Attention·needs·to·be·given·to·frequency·of·transfer,·and·whether·the·transfer·is·incremental·(i.e.·the·changes·between·snapshots)·or·a·complete·capture·of·content.··Frequency·of·transfer·would·be·determined·by·factors·such·as·when·data·is·deleted·or·overwritten·(e.g.·every·five·years),·or·when·there·is·significant·change,·for·example·an·annual·transfer·may·be·required·for·a·registration·database·that·is·updated·and·reviewed·at·the·end·of·the·financial·year.¶

Transfer Process Maps



Additional Metadata for Database/Business Systems

Purpose¶

This·document·identifies·additional·descriptive·metadata·elements·that·should·be·captured·to·better·describe·the·technical·structure·and·functionality·of·databases/business·systems.··While·the·CRS·system·adequately·captures·the·recordkeeping·context·of·record·creation·and·management,·digital·systems·require·at·least·some·of·their·technical·components·and·properties·to·be·identified·and·described·so·that·they·can·be·properly·understood·into·the·future,·long·after·they·have·ceased·to·exist.··This·metadata·is·also·important·when·providing·access·to·the·information·and·also·for·preservation·planning.··¶

Currently,·these·additional·elements·could·be·added·to·Series·and/or·Item·Notes·in·RecordSearch,·and·could·be·captured·in·discussions·with·agencies·when·gathering·information·about·the·database/business·system.··The·Series·Registration·Form·could·be·modified·to·include·these·elements.··In·the·fullness·of·time,·as·work·continues·on·the·National·Archives'·business·systems,·system·functionality·can·be·improved·to·better·accommodate·these·elements.¶

As·with·other·description·activity,·the·professional·judgement·of·the·transfer·officer/description·officer/archivist·should·be·used·when·determining·what·descriptive·information·to·capture.·This·document·is·provided·as·a·guide·and·prompt.··¶

The·Standards·Mapping·column·links·to·the·same·or·similar·properties·in·other·relevant·standards:·Archival·Control·Model·(ACM),·Data·Dictionary·for·Preservation·Metadata·(PREMIS),·Australian·Government·Recordkeeping·Metadata·Standard·(AGRkMS),·Software·Metadata·Recommended·Format·Guide·(SMRFG).¶

Metadata·Elements¶

Field¶ ¤	Description¤	Comment¤	Series·or·Item¤	Standards·mapping¤	¤
Database·Management·System¤	What·database·management·system·including·version·and·date,·was·used,·e.g.·SQL·Server,·MySQL,·Oracle,·PostgreSQL·etc?¤	Important·for·access·and·context.··Important·if·emulation·is·an·access·option.¤	Series·Note¤	ACM:·21.3·Creating·Application·Name¶ ACM:·21.4·Creating·Application·Version¶ PREMIS:·1.5.5.1·Creating·	¤

Project Outcome: Transferring Databases

- No one-size-fits-all approach
- Interpretation of the disposal class/analysis of business system
- Transfer could comprise a combination of:
  - SIARD file
  - Native SQL files
  - Export in an open data format like CSV, XML, etc
  - Transfer of reports/views
  - Documenting how the business system was used
  - Technical documentation – data dictionary, data model
  - Full descriptive information documented in the archival control system

Future Work

- Continue to develop knowledge and capability
- Pilot database transfers – Microsoft Dymanics, Lotus Notes
- Work with agencies to imbed transfer decisions in the Records Authority creation process

Thank You