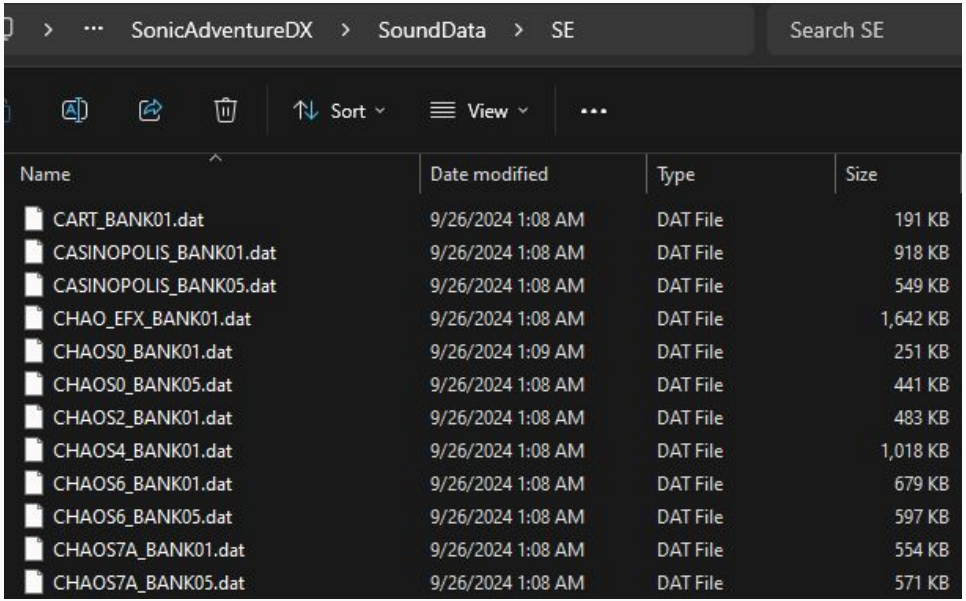
DATa Archive Tool

Documentation on the V2.DMZ format

Setup

- I will be using files from ‘Sonic Adventure DX’ Steam release for examples.
- If you have this game and want to follow along you can find them in \SoundData\SE\.
- For digging into the data of these files I will be using HxD, but any hex editor works



Getting Started

Let’s start with opening CASINOPOLIS_BANK05.dat

CASINOPOLIS_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	02	00	00	00	2C	00	00	00	00	10	00	00	00	40	04	00	....,.........@..
00000020	3A	00	00	00	01	50	04	00	00	40	04	00	42	30	35	5F	:....P...@..B05_
00000030	30	30	5F	30	30	2E	57	41	56	00	42	30	35	5F	30	30	00_00.WAV.B05_00
00000040	5F	30	32	2E	57	41	56	00	00	00	00	00	00	00	00	00	_02.WAV.........
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

Getting Started

This is a very simple file and only contains 2 WAV files.

CASINOPOLIS_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	02	00	00	00	2C	00	00	00	00	10	00	00	00	40	04	00	....,........@..
00000020	3A	00	00	00	01	50	04	00	00	40	04	00	42	30	35	5F	:....P...@..B05_
00000030	30	30	5F	30	30	2E	57	41	56	00	42	30	35	5F	30	30	00_00.WAV.B05_00
00000040	5F	30	32	2E	57	41	56	00	00	00	00	00	00	00	00	00	_02.WAV.........
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

Overview: *Header*

All of the files that are compatible with this method start with “archive V2.DMZ “

My program checks for this header before opening an archive.

Do not know why there is a double space between archive and V2, but the program keeps it.

CASINOPOLIS_BANK05.dat

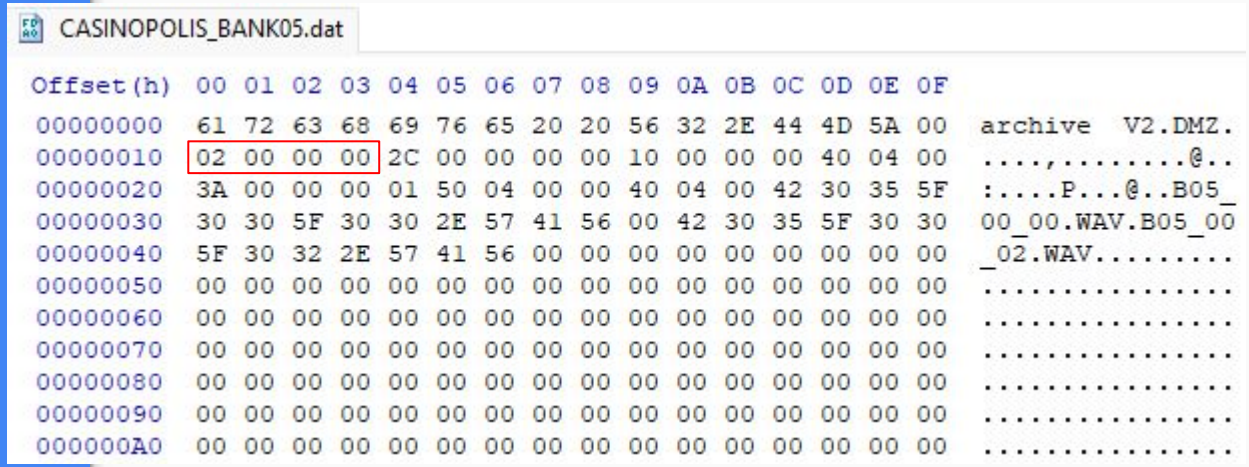
Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	02	00	00	00	2C	00	00	00	00	10	00	00	00	40	04	00	....,........@..
00000020	3A	00	00	00	01	50	04	00	00	40	04	00	42	30	35	5F	:....P...@..B05_
00000030	30	30	5F	30	30	2E	57	41	56	00	42	30	35	5F	30	30	00_00.WAV.B05_00
00000040	5F	30	32	2E	57	41	56	00	00	00	00	00	00	00	00	00	_02.WAV.........
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

Overview:
File Count

After the header, we go to **0x10** where this byte is used to keep track of how many files are stored in the archive.

The value is decimal converted to hex. So for this example 2=2, but a later example will show why this needs kept in mind.

It should be noted at this point too that the archive is formatted in words (4-byte chunks) and uses little-endian byte order.



Overview: *Index*

The space between the file count and the file names is used to store information for each file in the archive. Each file has three key values stored:

Word 1: Stores the offset of the filename in the archive.

Word 2: Stores the offset where the file data starts.

Word 3: Stores the size of the file in bytes.

CASINOPOLIS_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	02	00	00	00	2C	00	00	00	00	10	00	00	00	40	04	00	....,........@..
00000020	3A	00	00	00	01	50	04	00	00	40	04	00	42	30	35	5F	:....P...@..B05_
00000030	30	30	5F	30	30	2E	57	41	56	00	42	30	35	5F	30	30	00_00.WAV.B05_00
00000040	5F	30	32	2E	57	41	56	00	00	00	00	00	00	00	00	00	_02.WAV.........
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

Overview: *Filenames*

After the index is written the filenames are written sequentially

You can see here how the start of the filenames do match the offsets given

CASINOPOLIS_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	02	00	00	00	2C	00	00	00	00	10	00	00	00	40	04	00	....,........@..
00000020	3A	00	00	00	01	50	04	00	00	40	04	00	42	30	35	5F	:....P...@..B05_
00000030	30	30	5F	30	30	2E	57	41	56	00	42	30	35	5F	30	30	00_00.WAV.B05_00
00000040	5F	30	32	2E	57	41	56	00	00	00	00	00	00	00	00	00	_02.WAV.........
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

Overview:
File Data

Moving over to the next words for each file you can see where the offset is for the start of each of the files data

The first files data always starts at **0x1000**

Scrolling down to **0x1000** verifies we have the start of our data for the first file

CASINOPOLIS_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	02	00	00	00	2C	00	00	00	00	10	00	00	00	40	04	00	....,........@..
00000020	3A	00	00	00	01	50	04	00	00	40	04	00	42	30	35	5F	:....P...@..B05_
00000030	30	30	5F	30	30	2E	57	41	56	00	42	30	35	5F	30	30	00_00.WAV.B05_00
00000040	5F	30	32	2E	57	41	56	00	00	00	00	00	00	00	00	00	_02.WAV.........
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

CASINOPOLIS_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000F70	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000F80	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000F90	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000FA0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000FB0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000FC0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000FD0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000FE0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000FF0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00001000	80	00	01	78	03	12	04	02	00	00	56	22	00	03	C2	7F	€..x......V"..Â.
00001010	01	F4	04	00	00	00	00	00	00	00	00	00	00	00	00	00	.ô..............
00001020	00	0B	00	01	00	00	00	01	00	01	DB	20	00	02	18	00	..........Û ....

Overview:
File Length

The size of the file in bytes is stored. In the third word of the group.

Using this and the start offset for the data, programs reading the .dat file know where the data ends for a given file.

CASINOPOLIS_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	02	00	00	00	2C	00	00	00	00	10	00	00	00	40	04	00	....,.........@..
00000020	3A	00	00	00	01	50	04	00	00	40	04	00	42	30	35	5F	:....P...@..B05_
00000030	30	30	5F	30	30	2E	57	41	56	00	42	30	35	5F	30	30	00_00.WAV.B05_00
00000040	5F	30	32	2E	57	41	56	00	00	00	00	00	00	00	00	00	_02.WAV.........
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

Overview: Recap

Opening another .dat file with more files stored in it may better demonstrate the structure of this format.

Red, filename offsets

Blue, file data offsets

Green, file length

Comparing the files shows that the constants between all these .dat files is the header, the next word being used for file count, and that the first files data always starts at **0x1000**

CHAOVOICE_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	64	00	00	00	C4	04	00	00	00	10	00	00	CA	EA	00	00	d...Ä.....Êê..
00000020	D2	04	00	00	CB	FA	00	00	8A	7C	00	00	E0	04	00	00	Ò...Ëú..Š|..à...
00000030	56	77	01	00	9A	54	00	00	EE	04	00	00	F1	CB	01	00	Vw..šT..î...ñË..
00000040	5A	1C	00	00	FC	04	00	00	4C	E8	01	00	8A	85	00	00	Z...ü...Lè..Š…..
00000050	0A	05	00	00	D7	6D	02	00	76	5D	00	00	18	05	00	00	....×m..v]......
00000060	4E	CB	02	00	FE	29	00	00	26	05	00	00	4D	F5	02	00	NË..þ)..&...Mõ..
00000070	FE	56	00	00	34	05	00	00	4C	4C	03	00	C6	52	00	00	þV..4...LL..ÆR..
00000080	42	05	00	00	13	9F	03	00	A2	13	00	00	50	05	00	00	B....Ÿ..¢...P...
00000090	B6	B2	03	00	FE	0E	00	00	5E	05	00	00	B5	C1	03	00	¶²..þ...^...µÁ..
000000A0	1A	7D	00	00	6C	05	00	00	D0	3E	04	00	DA	71	00	00	.}..l...Ð>..Úq..
000000B0	7A	05	00	00	AB	B0	04	00	BE	6F	00	00	88	05	00	00	z...«°..¾o..ˆ...
000000C0	6A	20	05	00	1A	B3	00	00	96	05	00	00	85	D3	05	00	j ...³..–.....Ó..
000000D0	96	55	00	00	A4	05	00	00	1C	29	06	00	96	70	00	00	–U..¤....)..–p..
000000E0	B2	05	00	00	B3	99	06	00	6E	4D	00	00	C0	05	00	00	²...³™..nM..À...
000000F0	22	E7	06	00	BE	42	00	00	CE	05	00	00	E1	29	07	00	"ç..¾B..Î...á)..
00000100	AA	50	00	00	DC	05	00	00	8C	7A	07	00	16	36	00	00	ªP..Ü...Œz...6..
00000110	EA	05	00	00	A3	B0	07	00	36	37	00	00	F8	05	00	00	ê...£°..67..ø...

CHAOVOICE_BANK05.dat

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
000004B0	D6	E7	16	00	6E	17	00	00	2E	0A	00	00	45	FF	16	00	Öç..n.....Eÿ..
000004C0	9A	30	00	00	42	30	35	5F	30	30	5F	30	30	2E	57	41	š0..B05_00_00.WA
000004D0	56	00	42	30	35	5F	30	30	5F	30	31	2E	57	41	56	00	V.B05_00_01.WAV.
000004E0	42	30	35	5F	30	30	5F	30	32	2E	57	41	56	00	42	30	B05_00_02.WAV.B0

Overview: Recap

This file is also good for showing why it is important to remember the file count is a decimal as reading the hex value for what it is would have you believe the number of files is 64, however there are 100 files in this archive.

And that is all there is to this archive format, there is no extra data at the end of the archive so once the last files data is written the archive is fully saved.

CHAOVOICE_BANK05.dat

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	61	72	63	68	69	76	65	20	20	56	32	2E	44	4D	5A	00	archive V2.DMZ.
00000010	64	00	00	00	C4	04	00	00	00	10	00	00	CA	EA	00	00	d...Ä.......Êê..
00000020	D2	04	00	00	CB	FA	00	00	8A	7C	00	00	E0	04	00	00	Ò...Ëú..Š|..à...
00000030	56	77	01	00	9A	54	00	00	EE	04	00	00	F1	CB	01	00	Vw..šT..î...ñË..
00000040	5A	1C	00	00	FC	04	00	00	4C	E8	01	00	8A	85	00	00	Z...ü...Lè..Š…..
00000050	0A	05	00	00	D7	6D	02	00	76	5D	00	00	18	05	00	00	....×m..v]......
00000060	4E	CB	02	00	FE	29	00	00	26	05	00	00	4D	F5	02	00	NË..þ)..&...Mõ..
00000070	FE	56	00	00	34	05	00	00	4C	4C	03	00	C6	52	00	00	þV..4...LL..ÆR..
00000080	42	05	00	00	13	9F	03	00	A2	13	00	00	50	05	00	00	B....Ÿ..¢...P...
00000090	B6	B2	03	00	FE	0E	00	00	5E	05	00	00	B5	C1	03	00	¶²..þ...^...µÁ..
000000A0	1A	7D	00	00	6C	05	00	00	D0	3E	04	00	DA	71	00	00	.}..l...Ð>..Úq..
000000B0	7A	05	00	00	AB	B0	04	00	BE	6F	00	00	88	05	00	00	z...«°..¾o..ˆ...
000000C0	6A	20	05	00	1A	B3	00	00	96	05	00	00	85	D3	05	00	j ...³..–.....Ó..
000000D0	96	55	00	00	A4	05	00	00	1C	29	06	00	96	70	00	00	–U..¤....)..–p..
000000E0	B2	05	00	00	B3	99	06	00	6E	4D	00	00	C0	05	00	00	²...³™..nM..À...
000000F0	22	E7	06	00	BE	42	00	00	CE	05	00	00	E1	29	07	00	"ç..¾B..Î...á)..
00000100	AA	50	00	00	DC	05	00	00	8C	7A	07	00	16	36	00	00	ªP..Ü...Œz...6..
00000110	EA	05	00	00	A3	B0	07	00	36	37	00	00	F8	05	00	00	ê...£°..67..ø...
00000120	DA	E7	07	00	1E	6A	00	00	06	06	00	00	F9	51	08	00	Úç...j......ùQ..
00000130	B6	44	00	00	14	06	00	00	B0	96	08	00	FE	3B	00	00	¶D......°–..þ;..
00000140	22	06	00	00	AF	D2	08	00	3A	24	00	00	30	06	00	00	"...¯Ò..:$..0...
00000150	EA	F6	08	00	8E	21	00	00	3E	06	00	00	79	18	09	00	êö..Ž!..>...y...
00000160	26	17	00	00	4C	06	00	00	A0	2F	09	00	FE	3B	00	00	&...L... /..þ;..

End

<16 byte Header>

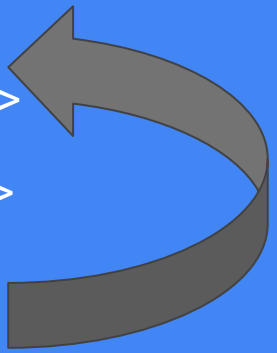
<4 bytes for # of files(Dec 2 Hex)>

<4 bytes for filename offset>

<4 bytes for file data offset>

First file data always at **0x1000**

<4 bytes for file length>



<Filenames>

<FileData>