Topic 3 - Introduction to the OAIS Reference Model

Speaker notes

No notes on this slide.

Open Archival Information System

ISO 147214:2012

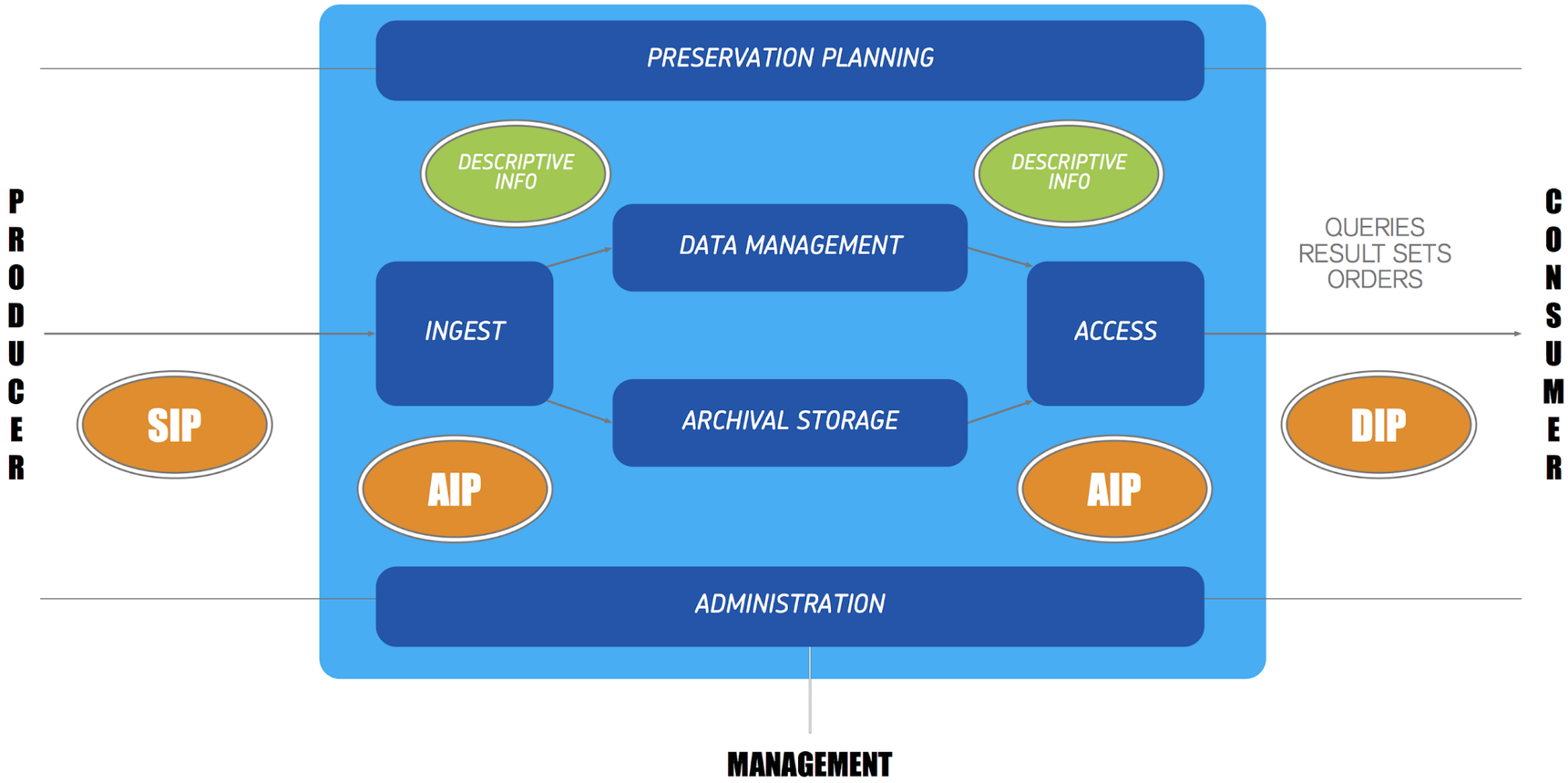
Speaker notes

Background and history:

NASA needed to come up with a good strategy for serious long-term preservation of their space flight data.

So they invented the OAIS model.

The Graph



Speaker notes

This is the basic OAIS model representation graph. You may encounter this graph in various different manifestations, but the components should always be the same.

Mandatory responsibilities

- Negotiate & accept appropriate information.
- Obtain sufficient control over the information.
- Determine scope of community needs.
- Ensure the preserved info is understandable independently without assistance.
- Safeguard your items (contingencies, deletions, …)
- **Make it available.**

Speaker notes

The reference model also defines mandatory responsibilities:

An OAIS-type archive must meet a set of six minimum responsibilities to do with the ingest, preservation, and dissemination of archived materials.

- From information producers: Negotiate for and accept appropriate information
- To meet long-term preservation objectives: Obtain sufficient control of the information
- Of the archive's user community: Determine its scope;
- Ensure that the preserved information is independently understandable to the user community. The information can be understood by users without the assistance of the information producer;
- Follow documented policies and procedures to ensure the information is preserved against all reasonable contingencies, and that there are no ad hoc deletions.
- Make the preserved information available to the user community,
  - Enable dissemination of authenticated copies of the preserved information…
  - In its original form or in a form traceable to the original.

The 3 Areas of Operation

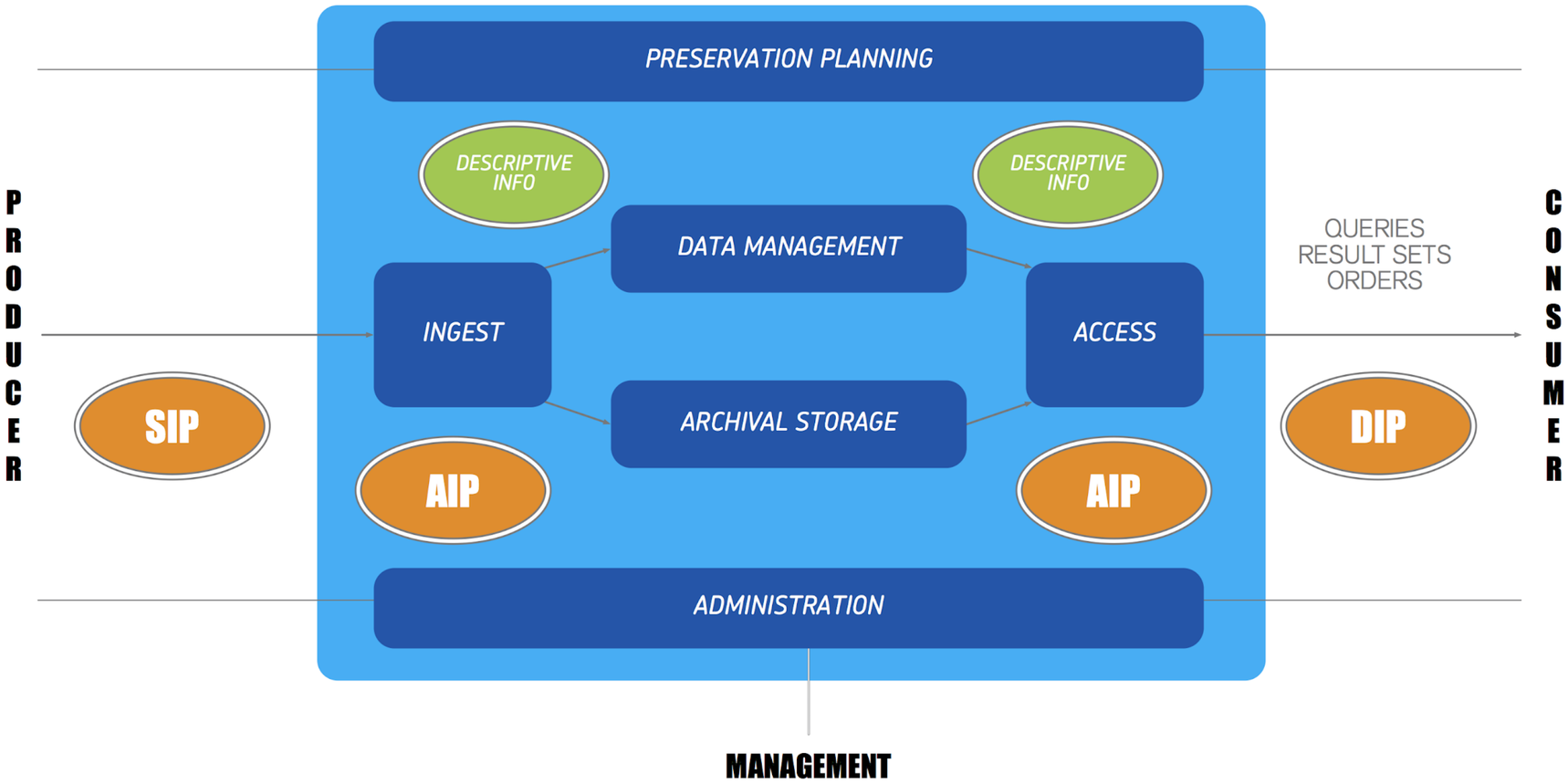
1. **External environment**
   producer, consumer, and management
2. **Functional components**
   of the repository itself
3. **Information packages**
   being preserved and disseminated

Speaker notes

The OAIS reference model describes three areas that collectively make up a repository's operation.

Each part plays equally important roles in the long-term conservation and access of digital information.

Locate the 3 Areas



Speaker notes

Give it a try :) We'll see the details later on.

- External environments: Producer, Consumer, Management
- Functional components: Preservation planning, Data Management, Archival Storage, Ingest, Access, Administration
- Information packages: SIP, AIP, DIP

Information Objects

"[are] either a physical or digital ***Data Object*** *with Representation Information that 'allows for the full interpretation of data into meaningful information'" —*

ImpactZone

Speaker notes

No notes on this slide.

Information Object Packages

Speaker notes

No notes on this slide.

SIP, AIP, DIP

- **SIP:**
  Submission Information Package
- **AIP:**
  Archival Information Package
- **DIP:**
  Dissemination Information Package

Speaker notes

What's a SIP?

A SIP is "what comes in" (from the producer), but it's not necessarily in a proper form to be preserved properly, or may not meet the archives requirements: so some actions are needed (metadata is added/extracted, a uniqute ID (Object ID, archive signature) is assigned, filename standardization/detox, etc).

There's more about different forms SIPs can have, later on in "Ingest".

What's an AIP?

The result of the afore mentioned preservation process is the "AIP". (usually pronounced: "Ey-Aye-Pee")

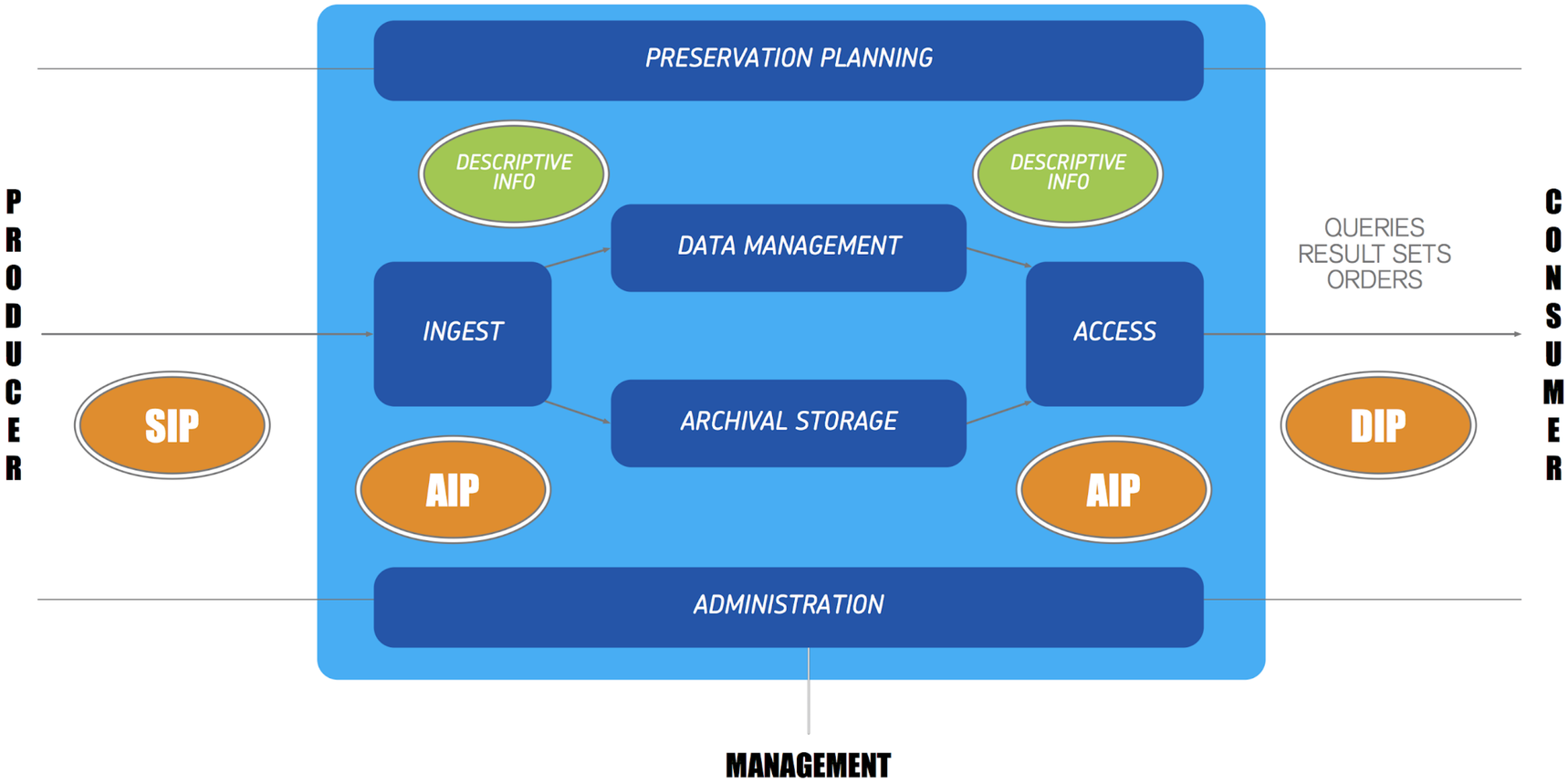
If done right, then the AIP can be used to generate any other derivative forms.

What's a DIP?

The DIP is "what goes out" (to the consumer), when someone requests a copy of the data stored as AIP. It might be the complete AIP as-is, but it could also just be a subset or even necessary/useful to put it in a different form (e.g. different codec/format).

Popular example for a DIP would be an access copy to put on a website for browsing the archive.

Find the Information Packages



Speaker notes

No notes on this slide.

Information Objects

"[are] either a physical or digital Data Object ***with Representation Information*** *that 'allows for the full interpretation of data into meaningful information'" —*

ImpactZone

Speaker notes

We've seen this definition before, but: What could be meant by "Representation Information" to be added to the Data Objects to form Information Objects?

Representation Information

Examples? Ideas?

Speaker notes

Hint: We've talked about how computer data can be represented, and what is needed to do so. Example: Text.

Representation Information

ASCII (1977/1986)

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
0_0	NUL 0000	SOH 0001	STX 0002	ETX 0003	EOT 0004	ENQ 0005	ACK 0006	BEL 0007	BS 0008	HT 0009	LF 000A	VT 000B	FF 000C	CR 000D	SO 000E	SI 000F
1_16	DLE 0010	DC1 0011	DC2 0012	DC3 0013	DC4 0014	NAK 0015	SYN 0016	ETB 0017	CAN 0018	EM 0019	SUB 001A	ESC 001B	FS 001C	GS 001D	RS 001E	US 001F
2_32	SP 0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
3_48	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
4_64	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
5_80	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
6_96	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
7_112	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	| 007C	} 007D	~ 007E	DEL 007F

Letter Number Punctuation Symbol Other undefined

Speaker notes

> *"Representation Information expressed in ASCII needs the additional Representation Information for ASCII, which might be a physical document giving the ASCII Standard" —* CCSDS, 2002

Information = a combination of Data and Information about how to represent that data => Representation Information.

The purpose of the Representation Information object is to convert the bit sequences into more meaningful information.

Representation Information

...for images, audio, video?

Speaker notes

With media files, that might be the information about how the bitstream of the audio/video codec is structured and therefore how it can be interpreted back as audio/video. Same goes for the container, etc.

This applies to all data formats, btw.

Representation Information

So, what could you store alongside your files?

Speaker notes

- Codec format specification (technical white paper)
- Encoding/decoding software
- Preview images in other formats or lower resolutions (to show how it should render)
- Example source code
- Pseudo code
- …

That’s why it’s so important to avoid undocumented (=black box) formats for preservation, as the Representation Information often cannot be supplied.

It’s a recursion.

Digital Representation Information is an Information Object that may have its own Data Object and its own Representation Information associated with understanding each Data Object.

Speaker notes

No notes on this slide.

...meaning:

- To preserve the meaning of an Information Object, its Representation Information must also be preserved.
- This recursion continues until physical forms, which can be understood by the Designated Community, are encountered.

Speaker notes

The good question is of course: What can you assume to be taken as “clear enough” so that you don’t have to provide it in the information package?

SIP/DIP: currently “well known” or widely supported “out of the box” stuff can be left out. AIP: hm… Rosetta stone?

Information Types

- Semantic
- Structure
- Other

Speaker notes

Information = a combination of Data and Representation Information

- Semantic Information: adds meaning to it. e.g. for a sequence of text characters the information as to which language was being expressed
- Structure Information: (often referred to as the ‘format’ of the digital object)
- Other Representation Information: e.g. Software needed to process a database file would be regarded as Other Representation Information.

Information Types

- **Semantic**
  Example: Text, but which language?
- **Structure**
  Example: File format definition (header, payload, …)
- **Other**
  None of the above ;)
  Example: Which software required

Speaker notes

No notes on this slide.

Common misunderstandings

- You don’t *have* to implement OAIS,
  but it’s a good reference.
- There’s no “common format” for SIPs, AIPs, DIPs.
  but there are good/best practices
- It doesn’t *have* to involve XML
  …or a specific metadata standard.
  (but it also doesn’t hurt ;P)
- Perfect can be the enemy of good
  (Aim for a sweet spot!)
- Don’t assume “*it’s clear to everyone*” WHO your designated users are.

Speaker notes

- Sweet spot = Minimalistic standard = As simple as possible, as complicated as necessary.
- Designated community/users: Take some time to discuss and maybe write it down, so it can be used to make informed decisions. Now and later on.

Comments?

Questions?

Speaker notes

No notes on this slide.

OAIS Background

- The Consultative Committee for Space Data Systems (CCSDS) initiated work aimed at developing formal standards for the long-term storage of digital data generated from space missions.
- The OAIS reference model was approved in January 2002 as ISO International Standard 14721;
- a revised and updated version was published in 2012 as ISO Standard 14721:2012.
- Design and revisions have taken place concurrent and in partnership with the growth of the profession into one that has boundaries, curricula, and standards of practice.

Speaker notes

No notes on this slide.

OAIS Background

- For a type of archive consisting of an organization of people and systems that has accepted the responsibility to preserve information for one or more designated communities.
- A reference model = an abstract framework or domain-specific ontology consisting of an interlinked set of clearly defined concepts produced by an expert or body of experts in order to encourage clear communication.
- Strictly a conceptual framework.
- Sets the standard for the activities that are involved in preserving a digital archive rather than the method for carrying out those activities

Speaker notes

No notes on this slide.

Links

- OAIS Reference Model (ImpactZone)
- The OAIS and the NSSDC (National Space Science Data Center)
- Preservation Metadata and the OAIS Information Model
- Open Archival Information System (Wikipedia)

Speaker notes

No notes on this slide.