



unIMC
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

i master
UNIVERSITÀ DI MACERATA

fgcAD
FORMAZIONE, GESTIONE E CONSERVAZIONE DI
ARCHIVI DIGITALI IN AMBITO PUBBLICO E PRIVATO

Il modulo ACCESS del sistema OAIS

Architetture e funzionalità avanzate

ASSITERM, 14-16 aprile 2016

Prof. Stefano Pigliapoco



unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Il Modello OAIS

- ✘ Il modello concettuale internazionale di riferimento per i sistemi di **Long-term digital preservation** è il modello OAIS (Reference Model for an Open Archival Information System), che è stato elaborato dal CCSDS (Consultative Committee for Space Data System) e successivamente approvato come standard ISO 14721:2012

Prof. Stefano Pigliapoco



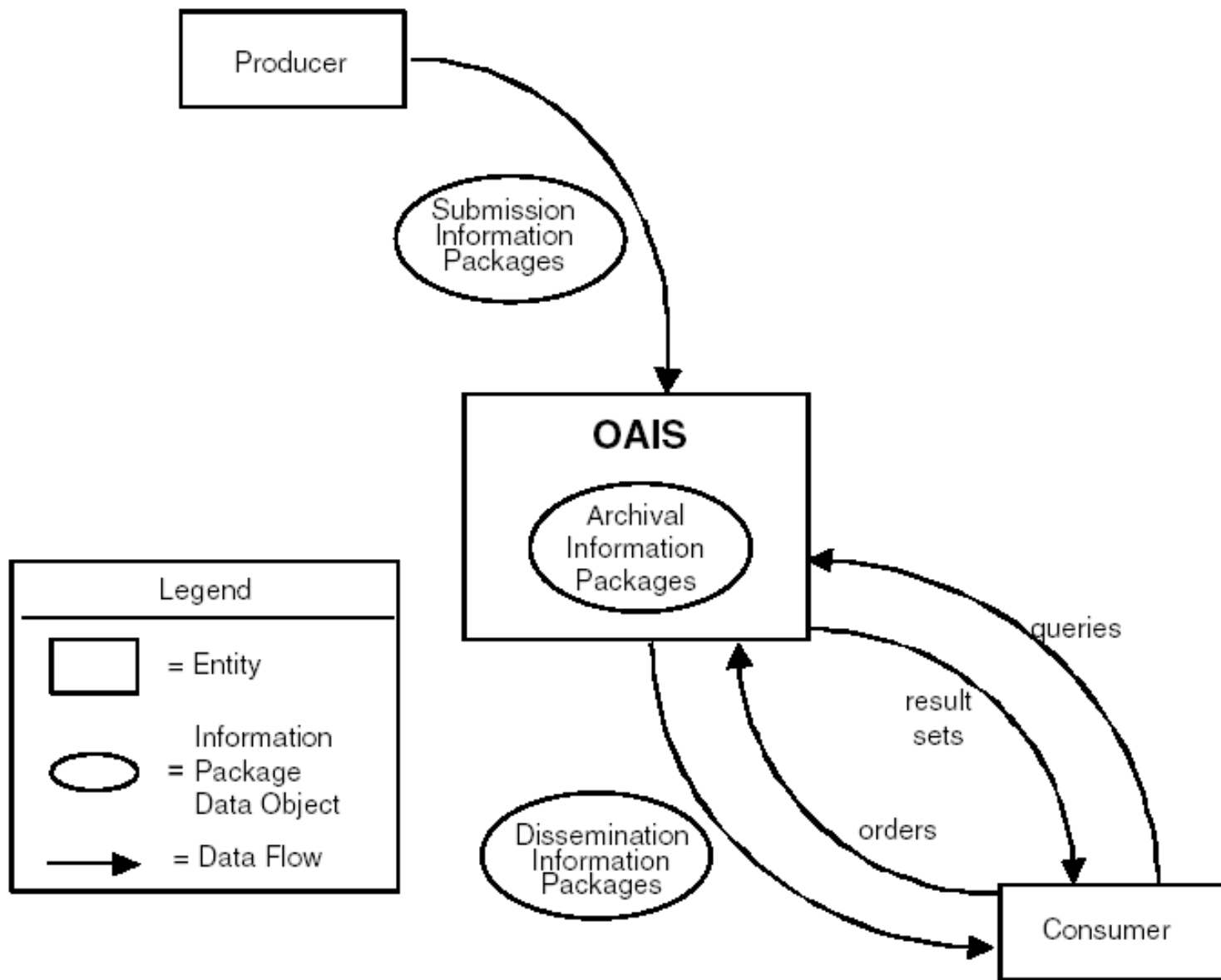
unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Il Modello OAIS

- ✘ Le nuove regole tecniche in materia di sistema di conservazione, emanate con DPCM 3 dicembre 2013, hanno **adeguato la normativa italiana agli standard internazionali** e adottato esplicitamente il modello concettuale OAIS per lo sviluppo dei servizi di *digital preservation*

Prof. Stefano Pigliapoco





unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

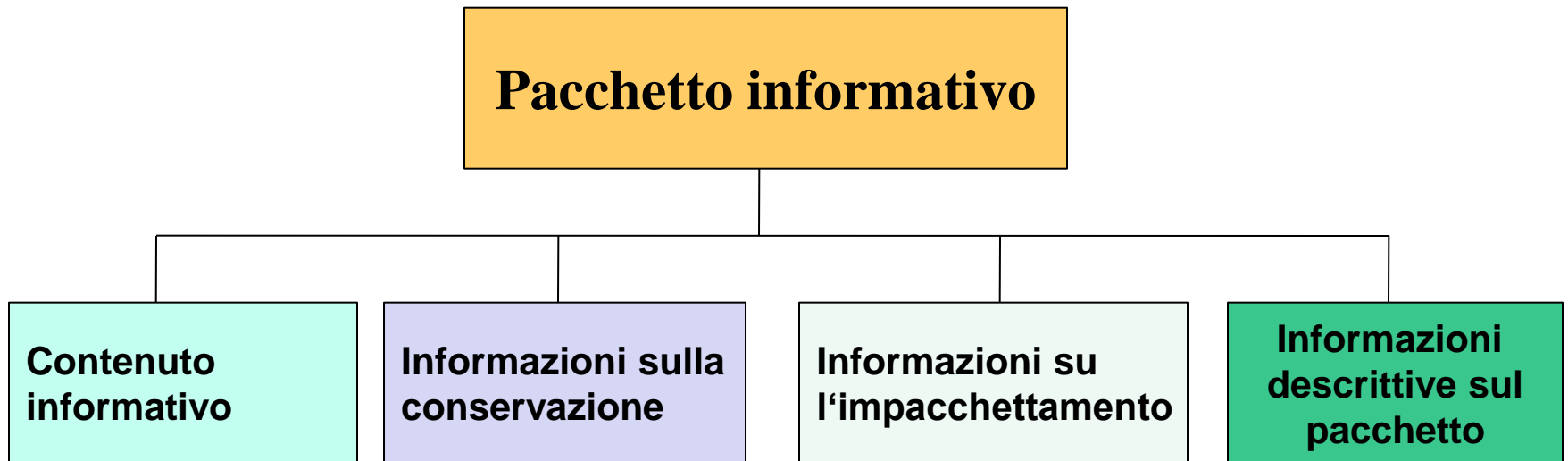
Pacchetti informativi

- ✘ La caratteristica più interessante dello standard ISO 14721:2012 è che i soggetti produttori, i conservatori e gli utenti possono scambiarsi solo **pacchetti informativi**, cioè strutture dati composte da quattro sezioni aventi ciascuna uno schema predefinito descritto nel modello OAIS

Prof. Stefano Pigliapoco



Pacchetti informativi





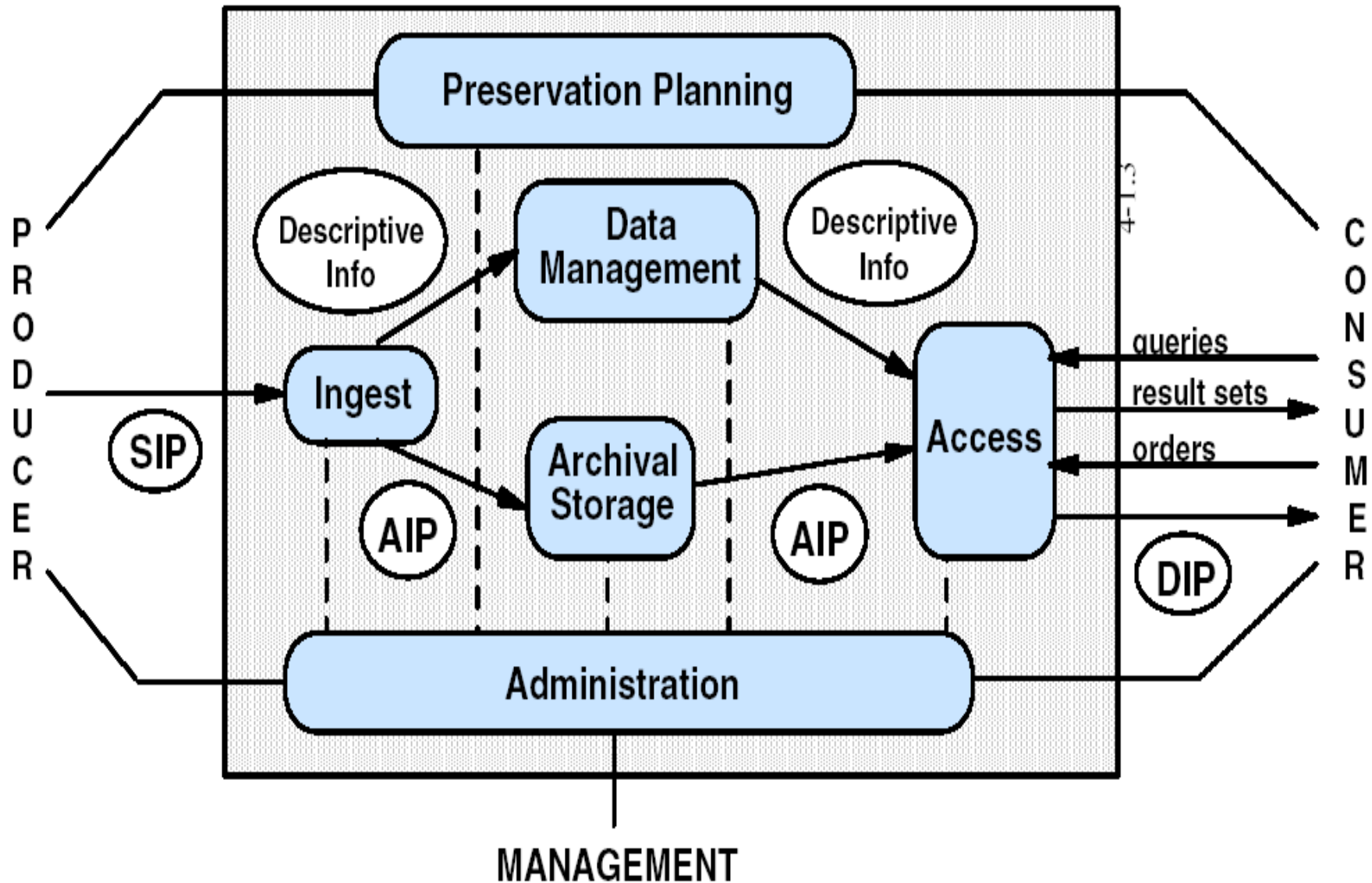
unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Pacchetti informativi

- ✘ L'uso dei pacchetti informativi equivale ad affermare un concetto importante:
 - il processo conservativo deve applicarsi ai contenuti digitali, ma anche **all'insieme delle informazioni** che li identificano, ne descrivono il contesto di produzione e di provenienza, li collegano ai precedenti e susseguenti.

Prof. Stefano Pigliapoco





unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Modulo ACCESS

- ✘ Il modulo di ACCESS realizza un'interfaccia attraverso la quale gli utenti possono inoltrare al sistema OAIS le loro **richieste di accesso**. Sono previste tre possibilità: a) consultazione di uno o più pacchetti informativi archiviati; b) richiesta di report di varia natura; c) richiesta di acquisizione di una o più entità conservate.

Prof. Stefano Pigliapoco



unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Accesso ai contenuti digitali

- ✘ I sistemi di conservazione basati sul modello OAIS si presentano come grandi depositi digitali alimentati continuamente, giorno dopo giorno, da una moltitudine di soggetti, pubblici e privati. La loro implementazione, pertanto, rappresenta un'opportunità da non perdere per attivare **funzionalità avanzate di l'accesso** ai contenuti digitali

Prof. Stefano Pigliapoco



unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Architetture tecnologiche

- ✘ In primo luogo occorre:
 - realizzare architetture informatiche che favoriscano la **condivisione** dei contenuti digitali attraverso i meccanismi **dell'interoperabilità e cooperazione applicativa**

Prof. Stefano Pigliapoco



unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Architetture tecnologiche

- ✘ Le architetture previste nello standard ISO 14721:2012 per l'interconnessione dei sistemi di conservazione digitale sono quattro:
- Independent archives
 - Cooperating archives
 - Federated archives
 - Shared resources archives

Prof. Stefano Pigliapoco



unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Architetture tecnologiche

- ✘ **Independent archives**: è la soluzione architeturale più semplice, che vede un sistema di conservazione offrire servizi a una comunità designata senza preoccuparsi di interagire e scambiare informazioni con altri sistemi. Di fatto è la soluzione adottata dai conservatori accreditati in Italia, i quali si preoccupano di acquisire i contenuti digitali **senza progettare forme di condivisione e collaborazione** con gli altri depositi

Prof. Stefano Pigliapoco



Architetture tecnologiche

- ✘ **Cooperating archives:** si basa sull'interoperabilità tra due o più sistemi di conservazione realizzata attraverso la condivisione dei modelli dei pacchetti informativi di versamento (SIP) e distribuzione (DIP)





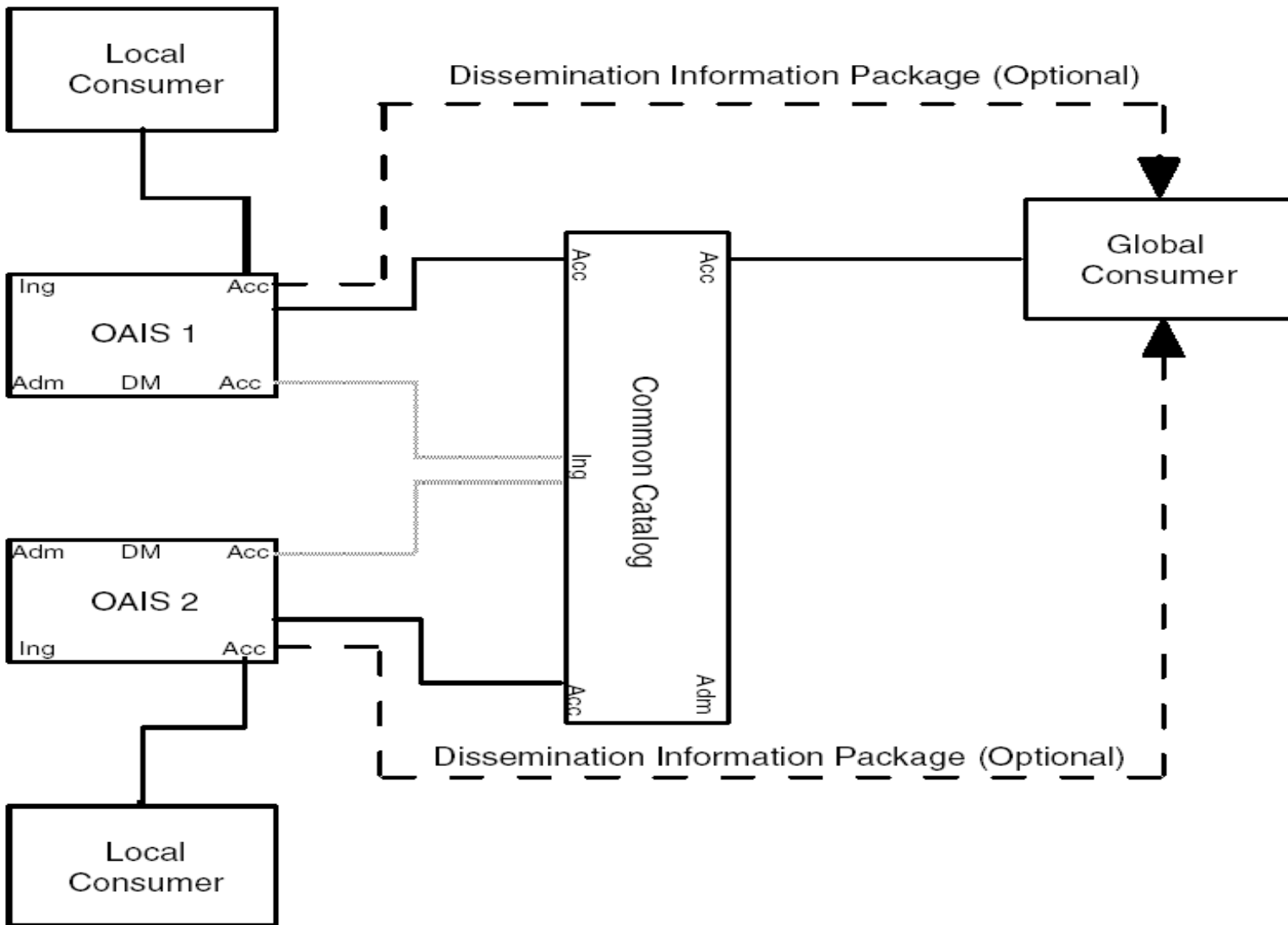
unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Architetture tecnologiche

- ✘ **Federated archives**: realizza un modello in cui più sistemi di conservazione indipendenti condividono una stessa interfaccia utente per consentire ad una comunità globale di consumatori di accedere a più depositi digitali con le stesse modalità. Tale soluzione richiede la costituzione di un **catalogo unico** del patrimonio informativo e documentario digitale e lo smistamento delle richieste degli utenti ai sistemi che conservano i contenuti desiderati

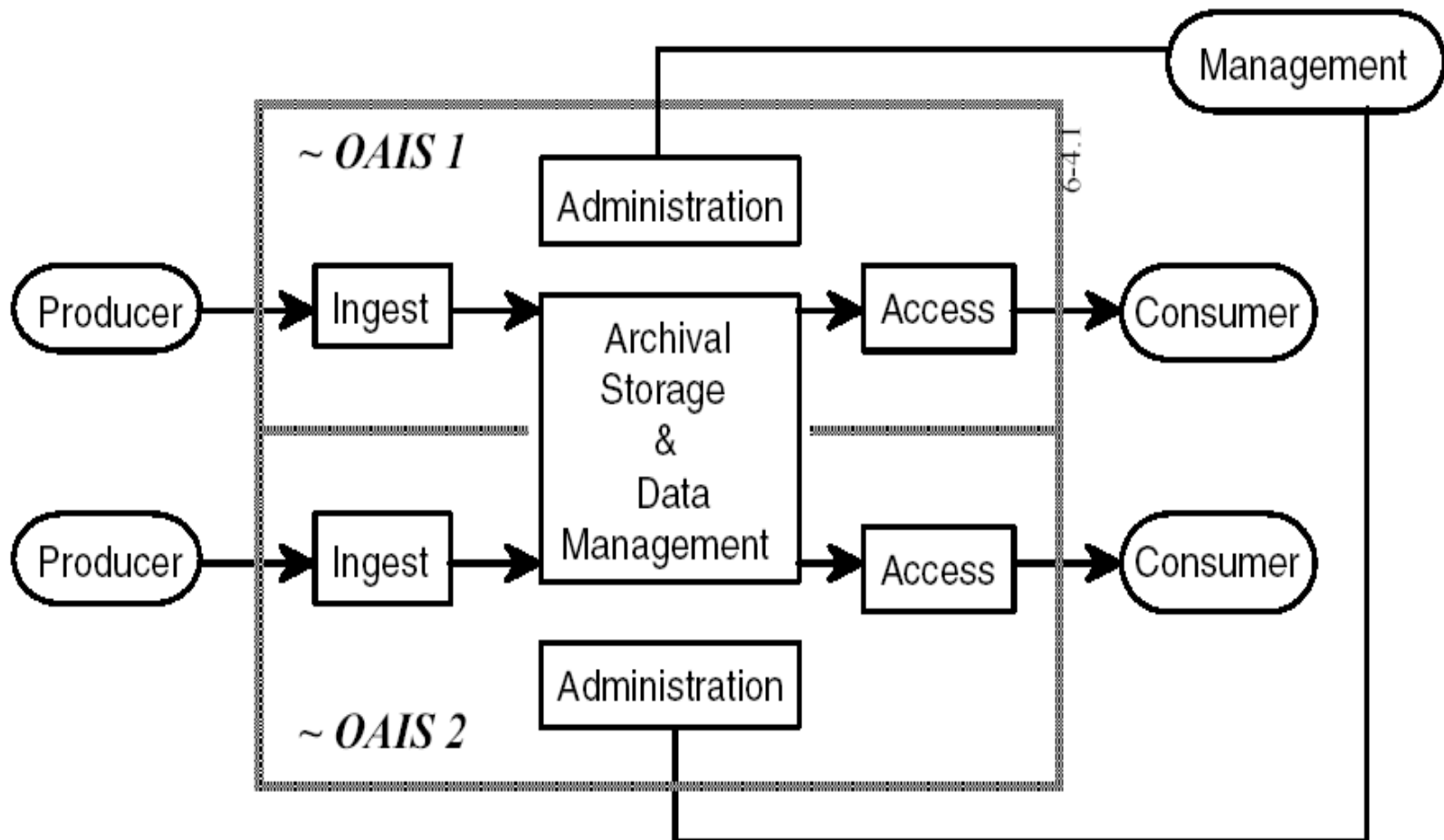
Prof. Stefano Pigliapoco





Architetture tecnologiche

- ✘ **Shared resources archives:** prevede la realizzazione di un deposito digitale in grado di interagire con più soggetti produttori e soddisfare le esigenze di accesso di più comunità di consumatori, grazie all'attivazione di componenti aggiuntive specializzate. L'utilizzo di questa architettura è giustificato dall'esigenza di ridurre i costi d'impianto condividendo uno o più moduli del sistema di conservazione





unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Standard di metadatazione

- ✘ Oltre che a livello tecnologico, per potenziare l'accesso ai contenuti digitali **occorre agire sui metadati** progettando attentamente i contenuti dei pacchetti informativi (SIP, AIP e DIP) con l'obiettivo di avere una base informativa sufficientemente ampia da soddisfare le esigenze di accesso di una comunità di utenti

Prof. Stefano Pigliapoco



Standard di metadattazione

- ✘ Esistono diversi standard di riferimento:
 - Dublin Core per i metadati descrittivi
 - METS - *Metadata Encoding and Transmission Standard* per i metadati amministrativi e gestionali, che può essere usato per creare i pacchetti informativi del modello OAIS
 - PREMIS - *Preservation Metadata: Implementation Strategies* per i metadati di conservazione



unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Standard di metadatazione

- ✘ L'Agenzia per l'Italia Digitale ha adottato per la struttura del pacchetto di archiviazione (AIP) lo schema definito nella norma UNI11386:2010 (UNI SinCRO) - *Supporto all'interoperabilità nella conservazione e nel recupero degli oggetti digitali*, che è lo standard nazionale a supporto del processo di conservazione

Prof. Stefano Pigliapoco



Standard di metadatazione

- ✘ Riuscire a generare per qualsiasi contenuto digitale il pacchetto AIP con la struttura dati UNI SinCRO, indipendentemente dagli standard di metadati descrittivi, amministrativi e strutturali utilizzati per rappresentarlo, non è affatto semplice. Si può agire sui valori *ExtraInfo*, ma questi dovrebbero essere **definiti formalmente per ogni tipologia di contenuto digitale e in funzione della comunità di riferimento**



Standard di metadatazione

- ✘ Allo stato attuale ci sono difficoltà sia presso i soggetti produttori sia nell'ambito delle strutture di conservazione. I primi non assicurano la valorizzazione di un adeguato set di metadati con la conseguenza che i SIP contengono poche informazioni utili in fase di ricerca. I conservatori non hanno progettato i loro sistemi con l'obiettivo di implementare architetture tecnologiche interoperabili né hanno ricercato l'uniformità nella definizione dei pacchetti informativi.



Motori di ricerca

- ✘ I motori di ricerca analizzano l'intero *world wide web* con l'obiettivo di individuare le pagine contenenti le sequenze binarie, corrispondenti a caratteri, simboli e parole chiave (*keywords*) specificate dall'utente e restituire i relativi *link*. Il compito di questi software è reso difficile dalla quantità enorme di documenti accessibili via web e dall'incapacità degli utenti non esperti di **esprimere compiutamente**, con una semplice breve frase, **le proprie esigenze informative**.



unimc
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Tecnologie semantiche

- ✘ La disponibilità in rete di una varietà di contenuti così ampia da sembrare illimitata unita all'esigenza di aumentare le potenzialità di accesso ha determinato lo sviluppo di nuove tecniche di ricerca basate sull'impiego delle **tecnologie statistiche, linguistiche e semantiche**

Prof. Stefano Pigliapoco



Tecnologie semantiche

- ✘ Queste tecniche presuppongono:
 - l'estrazione dei termini da un insieme di documenti
 - il controllo terminologico al fine di attribuire a ciascun termine il significato che ha in un dominio specialistico, evitando ambiguità e polisemia
 - l'elaborazione degli strumenti che costituiscono il *Knowledge Organization System* (KOS), rappresentati da tassonomie, thesaurus, mappe concettuali, ontologie



unIMC
UNIVERSITÀ DI MACERATA

l'umanesimo che innova

Tecnologie semantiche

- la rappresentazione formale degli strumenti KOS con un linguaggio *machine-readable* per dare ai sistemi la possibilità di utilizzare lo strato semantico e sviluppare i processi di “inferenza computazionale” andando oltre le tradizionali tecniche di *information retrieval*. I linguaggi comunemente utilizzati per la descrizione strutturata della conoscenza sono l’RDF e l’OWL

Prof. Stefano Pigliapoco



In conclusione

- ✘ Lo scenario che si profila all'orizzonte appare estremamente interessante riguardo alla capacità di accesso ai contenuti digitali in rete, ma per renderlo reale **occorre migliorare** i processi di produzione, archiviazione e conservazione dei contenuti digitali, senza dimenticare l'esigenza di formare **figure professionali** in grado di applicare concretamente, nei diversi contesti applicativi, le moderne tecnologie linguistiche e semantiche.