

**PROCEDURA APERTA PER L'ACQUISIZIONE DI SERVIZI DI SVILUPPO, MANUTENZIONE E
GESTIONE DELLA PIATTAFORMA PER LA CONSERVAZIONE A NORMA SACER**

**ALLEGATO 3A
CONTESTO APPLICATIVO E TECNOLOGICO**

Sommario

1. Contesto applicativo e tecnologico della Piattaforma	3
2. Contesto applicativo.....	3
2.1. Sistema di conservazione	3
2.1.1. Dimensionamento.....	3
2.1.2. Aspetti logico funzionali.....	3
2.1.3. Contesto applicativo.....	6
2.1.4. Sistema di gestione documentale	14
2.1.5. Sistema Sestra.....	16
3. Contesto Tecnologico	17
3.1. Infrastruttura	18
3.2. Linguaggi e Tecnologie	19
3.3. Framework Spagolite	20

1. Contesto applicativo e tecnologico della Piattaforma

Il presente allegato è dedicato alla descrizione del contesto applicativo e tecnologico della Piattaforma SACER e di alcuni dei sistemi ad essa collegati (Sestra e Sistema di gestione documentale).

2. Contesto applicativo

2.1. Sistema di conservazione

Il sistema Sacer è un sistema di conservazione a norma basato sul modello Open Archival Information System (OAIS - ISO 14721:2012) che svolge tutte le attività atte a proteggere e custodire nel lungo termine gli archivi di documenti e dati informatici garantendone l'autenticità, integrità, affidabilità, leggibilità e reperibilità.

Sacer supporta l'Amministrazione nello svolgimento delle funzioni di archiviazione e conservazione digitale dei documenti informatici sia per conto dell'Amministrazione stessa, sia per conto degli enti produttori, nel rispetto dei principi di efficacia, efficienza ed economicità, sulla base di quanto stabilito nelle vigenti Linee Guida AgID sulla formazione, gestione e conservazione dei documenti informatici al paragrafo 4.3.

Per quanto riguarda le Modalità organizzative con le quali l'Amministrazione eroga i servizi inclusi in questa richiesta si rimanda al capitolo 4 del Manuale di Conservazione, che illustra in dettaglio ruoli e responsabilità dei diversi attori del processo di conservazione realizzato dall'Amministrazione tramite il sistema di conservazione.

2.1.1. Dimensionamento

- numero enti produttori serviti: circa 1300
- numero utenti attivi: circa 4000
- numero sistemi versanti interfacciati: circa 130
- documenti annui versati (dato riferito al 2023): circa 314 milioni
- numero documenti conservati: circa 2,7 miliardi
- dimensione database: circa 90 TB
- dimensione oggetti conservati su Object Storage: circa 250 TB
- dimensione oggetti conservati su storage a cassette: circa 680 TB

2.1.2. Aspetti logico funzionali

Il contesto in cui è inserito l'applicativo presente in questo repository è quello del Sistema di Conservazione Sacer che svolge tutte le attività atte a proteggere e custodire nel tempo gli archivi di documenti e dati informatici garantendone l'autenticità, integrità, affidabilità, leggibilità e reperibilità.

Il diagramma in figura, realizzato sul modello della rappresentazione delle entità funzionali di **OAIS**, schematizza dal punto di vista logico le principali componenti del *Sistema di conservazione* di ParER e le principali relazioni con i soggetti interessati dal *processo di conservazione*.

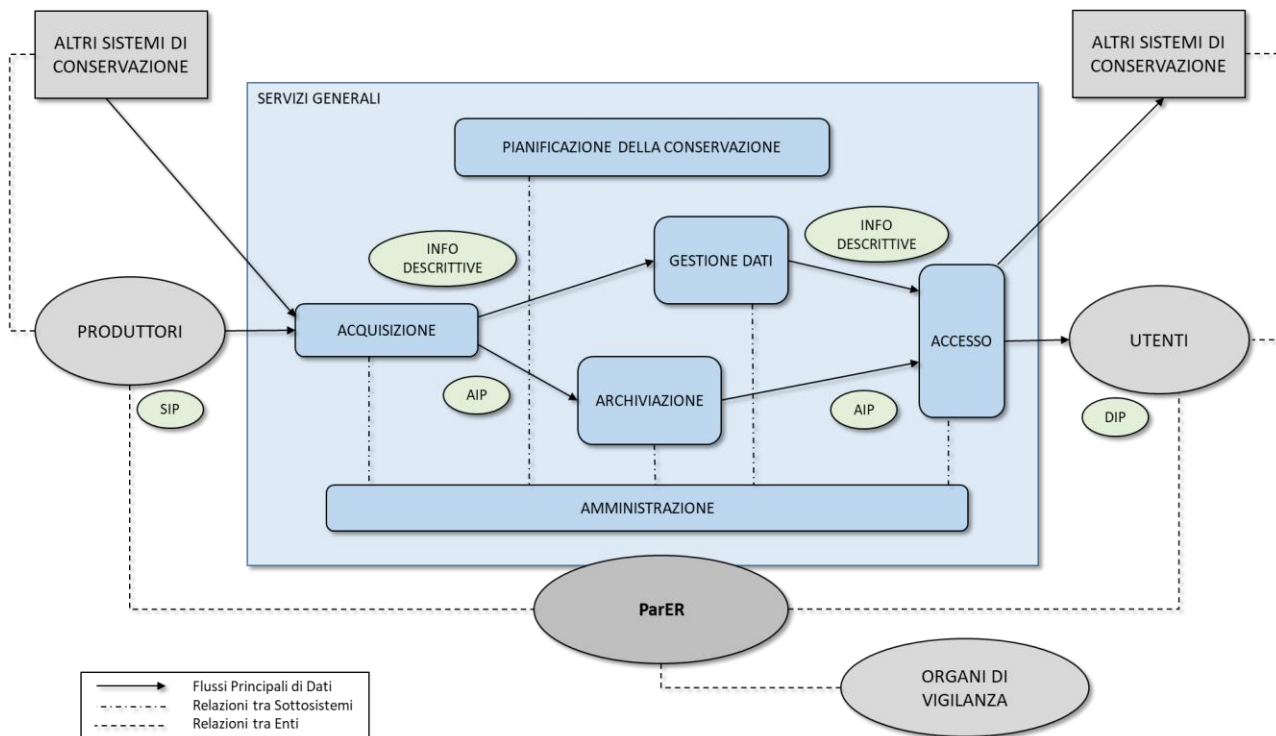


Figura 1 - Schema logico del Sistema di conservazione

Per la descrizione dei ruoli di *Produttori*, *Utenti*, *Regione Emilia-Romagna/ParER*, come soggetto Conservatore e Amministratore del Sistema, e *Organi di vigilanza* si rimanda al capitolo 4 del Manuale di conservazione¹.

Le funzionalità di **Acquisizione** gestiscono la fase di Acquisizione e presa in carico del processo di conservazione attraverso i **Web Service** di versamento esposti dal Sistema, consentendo la ricezione dei Pacchetti di versamento (SIP) dei *Produttori*, la loro verifica e la generazione, a partire dai SIP, dei relativi Pacchetti di archiviazione (AIP) e delle **Informazioni descrittive** per la loro ricerca.

In ottica di *interoperabilità* Sacer è in grado di ricevere AIP da altri *sistemi di conservazione*, di effettuare la migrazione in SIP e di versare tali SIP nel Sistema. Analogamente Sacer è in grado di trasmettere ad altri Sistemi gli AIP conservati, secondo gli accordi intercorsi con il *Produttore*.

Le funzionalità di **Gestione Dati** gestiscono le **Informazioni descrittive** generate al termine della fase di acquisizione e presa in carico del processo di conservazione. Tali funzionalità garantiscono: *memorizzazione*, manutenzione e aggiornamento all'interno del Sistema sia delle **Informazioni descrittive** necessarie a ricercare gli AIP, ricevute dall'Acquisizione, che dei dati necessari per gestire i pacchetti.

Le funzionalità di **Archiviazione** gestiscono la fase di gestione degli AIP del processo di conservazione: *memorizzazione*, **migrazione** dei supporti, backup, **Disaster recovery** ed eliminazione (*scarto*) degli AIP conservati nel Sistema.

1

Le funzionalità di **Amministrazione** gestiscono il governo dell'intero *processo di conservazione*, permettendo di definire e aggiornare nel Sistema politiche, standard e configurazioni che regolano tutte le altre funzionalità, incluse la gestione degli accordi con i *Produttori*, il monitoraggio del Sistema, la produzione di copie informatiche per la conservazione (**migrazione dei formati**) e la selezione degli AIP per lo *scarto*.

Le funzionalità di **Pianificazione** della conservazione gestiscono il monitoraggio dell'ambiente in cui il Sistema è inserito e forniscono le indicazioni necessarie per fare in modo che le informazioni conservate restino fruibili nel lungo periodo, tenendo conto dell'evoluzione tecnologica dei sistemi e del cambiamento della **Comunità di riferimento** (*Utenti*). Intervengono nella progettazione dei *Pacchetti Informativi* e nella pianificazione dello sviluppo e dei test del software necessario per la migrazione degli AIP. Tale funzione non è svolta da uno specifico applicativo, né segue procedure meccaniche, ma si configura invece come una serie di attività svolte utilizzando un insieme di strumenti, non solo informatici, finalizzati a raccogliere informazioni, confrontarsi con la **Comunità di riferimento**, effettuare test e verifiche sugli oggetti conservati, il tutto finalizzato a fornire indicazioni utili a mantenere il *processo di conservazione* aggiornato in relazione sia all'evoluzione tecnologica, che alle esigenze della **Comunità di riferimento**. I risultati di questa analisi si concretizzano, tipicamente ma non esclusivamente, in aggiornamenti nei modelli di *pacchetti informativi* gestiti dal Sistema, in implementazione di nuove librerie o altri strumenti software utilizzati dal Sistema, nella definizione e nell'aggiornamento delle politiche di conservazione e nei test su nuovi componenti hardware. Questi elementi sono inseriti nel Sistema utilizzando principalmente le funzionalità di Amministrazione di SacER e, secondariamente, quelle analoghe presenti negli altri moduli del Sistema, garantendo che il *processo di conservazione* sia sempre in grado tanto di contrastare efficacemente l'obsolescenza tecnologica, quanto di rispondere adeguatamente alle esigenze della **Comunità di riferimento** di ParER.

Le funzionalità di **accesso** gestiscono la fase di gestione del Pacchetto di distribuzione (DIP) del *processo di conservazione*, fornendo supporto agli operatori per la ricerca e la restituzione degli oggetti conservati. Le funzioni di *interoperabilità* consentono inoltre la restituzione da parte del Sistema di DIP coincidenti con gli AIP conformi a quanto previsto dalle Linee guida sulla formazione, gestione e conservazione dei documenti informatici² (d'ora in poi Linee Guida).

Il diagramma in figura schematizza i principali flussi di dati che intercorrono tra le componenti logiche del sistema descritte nei paragrafi precedenti; per completezza nello schema è stata inserita anche la componente '**Disaster recovery**', in quanto, pur non avendo un ruolo rilevante nella gestione ordinaria, riveste un ruolo significativo nello scambio di flussi informativi e svolge funzionalità elaborative autonome, seppur limitate, ai fini della produzione delle copie di salvataggio dei file su cassetta.

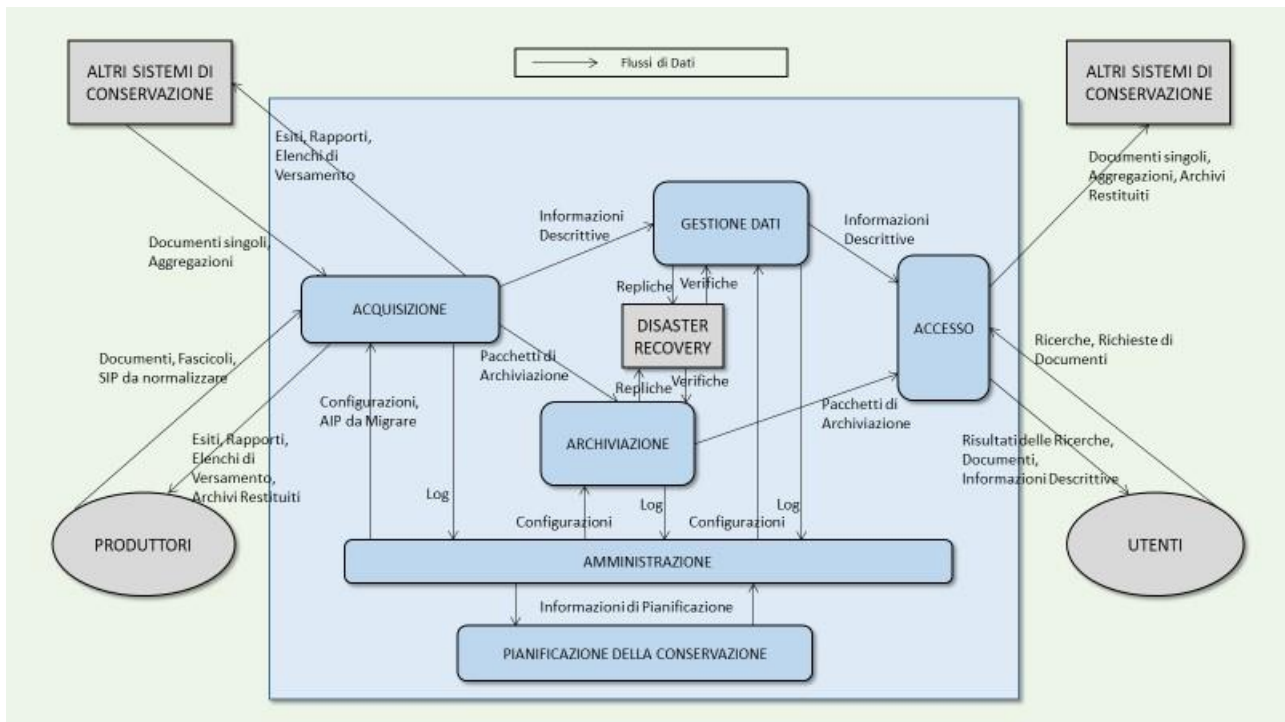


Figura 2 - Flussi di dati nel Sistema di conservazione

In aggiunta alle componenti logiche delineate nei paragrafi precedenti, che ne costituiscono il nucleo centrale, il Sistema mette a disposizione diversi Servizi generali a supporto delle altre funzionalità. Oltre ai servizi di gestione e monitoraggio dei sistemi, della rete, e della sicurezza dei sistemi, mette a disposizione in particolare:

- il servizio di **Identity Management**, che, via *IDP*, garantisce i corretti accessi al Sistema da parte dei diversi utilizzatori. I Sistemi consentono l'accesso anche via *SPID*;
- il servizio di Firma tramite dispositivo **HSM**, che consente di apporre le firme digitali necessarie nel processo di conservazione;
- il servizio di **Audit e Log**, che mantiene e manda in conservazione la storia degli accessi effettuati al Sistema e ai dati.

Per quanto riguarda il Quadro funzionale attuale del sistema di conservazione si rimanda al capitolo 8.1 del Manuale di Conservazione, che illustra in dettaglio le funzionalità del sistema di conservazione in coerenza con lo standard OAIS.

2.1.3. Contesto applicativo

Il *Sistema di conservazione* è costituito da diversi moduli software che interagiscono tra loro per la gestione dell'intero *processo di conservazione*. Il Sistema, inoltre, si avvale di ulteriori componenti applicative esterne con funzioni di supporto al processo.

Il diagramma in figura schematizza dal punto di vista tecnologico le principali componenti del *Sistema di conservazione* di ParER e le principali relazioni con altri sistemi interessati dal *processo di conservazione*.

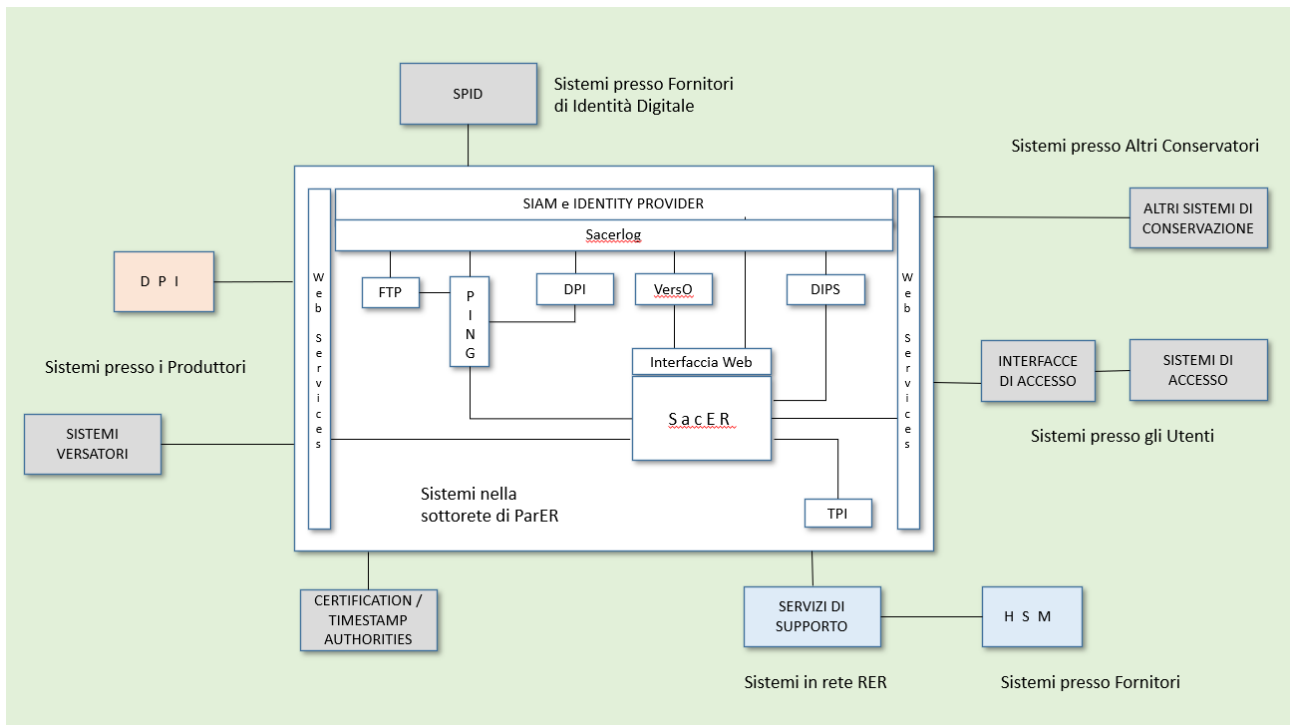


Figura 3 - Schema Tecnologico del Sistema di conservazione

Nella figura sono riportate:

- in bianco le componenti del Sistema di conservazione interne al perimetro di sicurezza;
- in rosa le componenti del Sistema di conservazione sviluppate da ParER, ma esterne al perimetro di sicurezza del data center di ParER, in quanto installate nelle reti dei Produttori;
- in azzurro le componenti a supporto del Sistema gestite dalla Regione Emilia-Romagna, anch'esse interne al perimetro di sicurezza;
- in grigio le componenti che fanno riferimento ai soggetti esterni (Produttori, Utenti e Authorities).

Per quanto riguarda il Contesto operativo di riferimento in cui si inseriscono i sistemi applicativi interessati dai servizi richiesti si rimanda al capitolo 7 del Manuale di Conservazione, che illustra in dettaglio il processo di conservazione realizzato dal ParER tramite il sistema di conservazione.

Qui di seguito sono illustrati i principali moduli software del Sistema e le componenti di supporto.

2.1.3.1. SacER

Il modulo software SacER costituisce il nucleo centrale del Sistema e implementa le funzionalità principali del *processo di conservazione*:

- Acquisizione;
- Gestione dati;
- Archiviazione;
- Amministrazione;
- Accesso.

Le funzionalità di **Acquisizione** di SacER consentono la gestione delle varie fasi del processo di Acquisizione e *presa in carico*. Operativamente si compongono delle seguenti attività:

- acquisizione del SIP trasmesso attraverso i Servizi di versamento;
- **memorizzazione** del SIP in un'area temporanea logicamente distinta dall'*archivio* vero e proprio per effettuare le verifiche previste;
- verifica del SIP in relazione alla struttura dati, ai *metadati* degli oggetti sottoposti a conservazione, alle eventuali firme apposte sui file (**Oggetti-dati**) associati ai **Componenti**, ai *formati* dei file stessi e generazione del *Rapporto di versamento* in caso di verifica positiva;
- restituzione dell'**Esito versamento**, comprensivo, in caso di esito positivo, del *Rapporto di versamento*;
- creazione degli **Elenchi di versamento**: attraverso vari job eseguiti sugli Application server automaticamente e in modo ricorrente secondo intervalli di tempo stabiliti nel modulo Amministrazione i SIP sono raggruppati in Elenchi di versamento, validati dal Responsabile della funzione ed elaborati per produrre i relativi AIP;
- eventuale migrazione di **formato** degli **Oggetti-dati** contenuti nei SIP sia per contrastarne l'obsolescenza tecnologica, sia per esigenze di miglioramento della fruibilità degli oggetti conservati;
- estrazione dei metadati dal SIP (ed eventuale loro normalizzazione) e dal Sistema da utilizzare per completare le informazioni necessarie a generare l'AIP (**Informazioni sulla rappresentazione, Informazioni sulla conservazione, Informazioni sull'impacchettamento, Informazioni descrittive**);
- generazione dell'**Indice dell'AIP**, utilizzando i metadati estratti dal SIP e quelli generati dal Sistema nel corso del processo di conservazione. SacER produce report di eccezioni a fronte di situazioni anomale nella creazione dell'**Indice dell'AIP**. Tutte le azioni vengono registrate sul sistema in apposite tabelle di log;
- generazione dell'AIP, che avviene impacchettando in un oggetto auto-consistente l'**Indice dell'AIP**, le evidenze informatiche prodotte nel corso del processo di conservazione e gli **Oggetti-dati**.

Le funzionalità di **Gestione Dati** di SacER sono finalizzate principalmente a gestire le **Informazioni descrittive** degli AIP generate durante il processo di acquisizione e includono:

- la **memorizzazione** dei *metadati* estratti dal SIP o generati dal Sistema nel corso del processo di Acquisizione dei SIP;
- la gestione degli aggiornamenti dei dati generati dalle funzionalità di Amministrazione e nel corso del *processo di conservazione*;
- l'esecuzione delle ricerche e la sua restituzione alle funzionalità di Accesso, che avvengono mediante l'utilizzo di funzionalità da interfaccia web del Sistema o mediante chiamata a Servizi specifici.

Le funzionalità di **Archiviazione** di SacER gestiscono la parte del processo di Gestione dell'AIP relativa alla *memorizzazione* e verifica degli **Oggetti-dati** su Data Base Oracle e **file system** (vedi paragrafo 7.5) e comprendono:

- la **memorizzazione** degli AIP e l'organizzazione gerarchica dei supporti di memorizzazione. In particolare, gli **Oggetti-dati** degli AIP, costituiti dagli **Indici degli AIP** e dei SIP, dagli **Esiti versamento**, dai *Rapporti di versamento* e dai file associati ai **Componenti**, sono memorizzati su supporti di diverso tipo in ragione della loro dimensione e della frequenza con cui vengono ricercati:
 - gli **Oggetti-dati** vengono memorizzati in modo permanente nell'**object storage**;
 - gli **Oggetti-dati** di grande dimensione e di accesso meno frequente vengono salvati temporaneamente su **file system** in cartelle opportunamente strutturate, per poi essere memorizzati in modo permanente su supporti a cassette, utilizzando a questo scopo un apposito componente software chiamato TPI, descritto più avanti;

- il controllo dell'integrità degli oggetti conservati, comprensivo della copia degli archivi, del controllo degli errori e delle procedure di refreshing dei supporti, in conformità con il Piano della Sicurezza del servizio di conservazione;
- la restituzione dei pacchetti alle funzioni di Accesso, mediante opportune funzionalità dell'interfaccia web del Sistema o mediante l'utilizzo di Servizi di recupero esposti da Sacer e descritti nel documento Specifiche tecniche dei servizi di recupero;
- l'annullamento dei versamenti effettuati, mediante opportune funzionalità dell'interfaccia web del Sistema o mediante l'utilizzo di Servizi di annullamento esposti da Sacer e descritti nel documento Specifiche tecniche dei servizi di annullamento;
- la cancellazione degli AIP sottoposti a scarto o a seguito di restituzione dei documenti conservati ai Produttore alla cessazione del servizio.

Le funzionalità di **Amministrazione** di SacER consentono di gestire configurazioni e parametrizzazioni in grado di determinare il funzionamento del Sistema in funzione degli specifici accordi intercorsi con i *Produttori*. Inoltre, consentono di monitorare tutta l'attività svolta da SacER. In particolare, in SacER è possibile configurare tutte le entità significative: Enti, **Strutture**, operatori e relativi profili, **tipologie documentarie**, **formati** accettati, logiche di controllo dei versamenti, logiche di creazione delle **Serie**, regole di *accesso* e di esibizione, politiche di monitoraggio del sistema. Anche l'interfaccia web di SacER è configurata automaticamente in ragione del profilo dei singoli operatori che vi accedono.

Le funzionalità di Amministrazione sono costituite da transazioni eseguibili tramite l'interfaccia web del sistema e riservate agli operatori di ParER, ma visibili negli esiti anche agli operatori dei *Produttori*.

Le funzionalità di **Accesso** di SacER consentono di restituire in forma di DIP gli oggetti conservati. A tal fine SacER mette a disposizione un'interfaccia web per le ricerche e per l'estrazione manuale dei documenti e dei Servizi di recupero per l'estrazione automatica.

Gli AIP forniti sono trasformati in DIP sulla base delle caratteristiche dell'oggetto e degli utilizzi cui è destinato. In molti casi la trasformazione dell'AIP in DIP può richiedere specifici passi elaborativi e trasformazioni complesse, che, necessitando di elaborazioni onerose, vengono normalmente eseguite da opportuni job batch e mantenute in modo permanente sul Data Base. Secondo la natura dei DIP, l'*esibizione* può avvenire on-line con un download, oppure tramite il trasferimento in un'area di transito, da cui il successivo recupero viene effettuato dal sistema richiedente con chiamata FTP. In molti casi, per comodità di trasferimento e recupero, i vari elementi che costituiscono il DIP vengono compressi in un archivio di tipo ZIP.

Il modulo di Accesso, oltre a verificare tramite i servizi di Autenticazione l'abilitazione dell'*Utente* al recupero del documento, traccia in apposite tabelle di log tutte le richieste prevenute, qualunque ne sia stato l'esito.

2.1.3.2. PING

Il modulo software PING (PreINGest) gestisce il processo di preacquisizione nel caso di **versamento** di Oggetti da trasformare in SIP mediante processi ETL.

La trasmissione dei pacchetti, solitamente compressi, avviene tramite interfaccia utente o tramite protocollo **FTPS**; l'**FTP server** provvede a memorizzare i file ricevuti sullo **storage** dedicato allo spazio FTP di input. Una volta ricevuti gli Oggetti, un job schedulato provvede alla loro trasformazione in SIP da versare a Sacer mediante processi ETL gestiti dal modulo Kettle-server descritto più avanti. Un ulteriore job schedulato si occupa di effettuare il **versamento** a SacER, che avviene utilizzando un apposito servizio di versamento. Tale servizio accetta in chiamata due file XML, uno con l'**Indice del SIP** e un altro con le **Informazioni sull'impacchettamento**, relative alla posizione dei file del SIP memorizzati sullo spazio FTP di input.

SacER utilizza le **Informazioni sull'impacchettamento** per recuperare i file dal **file system** di PING e depositarli nel proprio per le successive elaborazioni.

Le successive elaborazioni vengono eseguite da PING direttamente, nel caso in cui la normalizzazione possa basarsi su regole precodificate (come p.e. nel caso delle immagini diagnostiche in formato Dicom), oppure utilizzando un motore ETL di esecuzione delle trasformazioni nel caso in cui si debbano applicare regole di trasformazioni specifiche dell'oggetto in questione. In questo caso le regole vengono definite durante la fase di avvio del servizio, tramite uno strumento visuale ed eventuali integrazioni di codice sviluppato ad hoc.

PING traccia e memorizza nel proprio Data Base gli esiti dei versamenti a SacER e può essere interrogato da un operatore tramite interfaccia web o dal sistema versante tramite opportuno **Web Service**, per conoscere a quale punto del processo è giunto il SIP.

PING mette inoltre a disposizione del *Produttore* un client di versamento di Oggetti da trasformare, sia on line, sia tramite l'utilizzo di un client **FTP** installato sulla postazione di lavoro dell'utente o su un server della rete del *Produttore*.

Kettle Server

Il modulo Kettle Server espone, tramite un server Tomcat, delle API di tipo SOAP e REST che, su richiesta di PING, permettono l'esecuzione delle trasformazioni necessarie a generare il SIP a partire dall'oggetto da trasformare.

Le trasformazioni sono realizzate ed eseguite con un motore ETL, basato sulle librerie di Pentaho Data Integration (aka Kettle)³, inglobato nel modulo applicativo.

Kettle server fa da interfaccia tra l'utente esterno e il database delle trasformazioni, si occupa di preparare l'ambiente per la trasformazione, definire i parametri specifici per l'esecuzione della trasformazione.

Kettle server si occupa anche del monitoraggio e della gestione dello storico delle trasformazioni eseguite.

2.1.3.3. VersO

Il client di versamento manuale VersO (Versamento Online) è un modulo a disposizione degli Enti produttori per svolgere operazioni routinarie sul sistema. Poiché utilizza un'interfaccia web, non richiede l'installazione di alcun software sulla stazione di lavoro dell'Utente.

Il suo utilizzo tipico è il **versamento di Unità documentarie** per le quali non esiste un sistema interfacciato con Sacer. VersO viene richiamato tramite interfaccia web, si autentica sull'**IdP** di ParER o su **SPID**, utilizzando in ogni caso logiche di profilazione del Sistema, ed effettua il **versamento** dei SIP tramite interazione guidata con l'operatore del *Produttore*.

Tale modulo semplifica le operazioni di **versamento** manuale da parte del *Produttore*, automatizzando la generazione dell'**Indice del SIP** ed effettuando un test completo della correttezza del versamento prima di eseguire il versamento stesso.

Inoltre, mantiene il log dei versamenti effettuati e consente di interrompere temporaneamente l'operazione (p.e per raccogliere informazioni necessarie per completarlo), riprendendola successivamente, indipendentemente dalla scadenza della sessione web.

³ <https://www.hitachivantara.com/pentaho/pentaho-plus-platform/data-integration-analytics/download-pentaho.html>

2.1.3.4. DPI

Il modulo software DPI (Digital Preservation Interface), sviluppato e mantenuto dall'Amministrazione, consiste in un sistema di interfaccia tra i sistemi dell'Ente produttore e PING. DPI può essere installato all'interno della rete dell'Ente stesso e gestito secondo le politiche di sicurezza dell'Amministrazione, potendo tra l'altro autenticarsi sul suo **IdP**, oppure installato presso il Produttore.

DPI implementa funzionalità di **versamento** per specifiche tipologie di SIP. In particolare, qualificandosi come nodo **DICOM**, DPI riceve dai **PACS** studi diagnostici, che poi trasmette a PING per la trasformazione e il **versamento** a SacER.

DPI può operare con logiche sia push che pull, ricevendo o estraendo dati e documenti dai sistemi del *Produttore* per poi versarli nel Sistema, richiamando gli opportuni servizi di PING.

Inoltre, DPI fornisce strumenti di monitoraggio dei versamenti effettuati a disposizione dell'Ente produttore.

2.1.3.5. Interfacce (API)

SACER espone API REST e SOAP per dialogare sia con i sistemi interni che tra i diversi moduli applicativi. Le API possono essere esposte da moduli applicativi (SACER, PING), essere raccolte in un contesto applicativo specifico (SacerWS, vedi paragrafo successivo) o sviluppate come microservizi in un'architettura service oriented (vedi paragrafo 0).

SacerWS

SACERWS è il modulo del Sistema di Conservazione SACER che espone i principali web service di versamento, aggiornamento e recupero degli oggetti conservati.

I servizi, descritti nei documenti "Specifiche tecniche dei servizi di versamento", Specifiche tecniche dei servizi di versamento dei fascicoli, Specifiche tecniche dei servizi di aggiornamento dei metadati, sono i seguenti:

- Versamento Unità documentaria
- Aggiunta documento
- Aggiornamento Unità Documentaria
- Versamento Fascicolo
- Versamento MultiMedia (non esposto all'esterno)

Microservizi

Numerose API esposte dal Sistema, soprattutto per gestire processi e attività interni al Sistema, sono state sviluppate secondo il modello architetturale a microservizi. A livello funzionale, le più rilevanti per utilizzo e/o impatto per l'esterno sono:

- **Verifica Firma Eidas⁴**
Effettua la verifica e la validazione di firme digitali conformi allo standard eIDAS. Non è esposto all'esterno, viene invocato da SacerWS nell'ambito delle verifiche effettuate al versamento e aggiornamento dei SIP.
È realizzato attraverso framework [Spring Boot](#) (versione 3.x) e [OpenJDK 17](#), utilizza la libreria del progetto [DSS](#).

⁴ https://developers.italia.it/it/software/r_emiro-regioneer-parer-verifica-firma-eidas

- **Verifica Firma Crypto⁵**

Effettua la verifica e la validazione di firme meno recenti con formato non conforme allo standard eIDAS. Non è esposto all'esterno, viene invocato da SacerWS nell'ambito delle verifiche effettuate al versamento e aggiornamento dei SIP.

È realizzato attraverso framework [Spring Boot](#) (versione 3.x) e [OpenJDK 17](#).

- **Versamento Fascicolo**

Consente a un sistema versante esterno di versare a Sacer un fascicolo.

2.1.3.6. TPI

Il modulo software TPI (Tivoli Preservation Interface) gestisce la *memorizzazione* degli **Oggetti-dati** su supporti a cassette, operata utilizzando come sistema di gestione della **tape library** il software Tivoli.

TPI opera nel seguente modo:

- un job schedulato sul file server invia al sistema di gestione della **tape library** il comando di archiviazione delle cartelle in cui SacER ha depositato gli oggetti da archiviare, selezionate tramite opportuni criteri definiti in sede di amministrazione di sistema;
- il sistema di gestione della **tape library** provvede a leggere i file dalle cartelle e ad archivarli tramite le sue funzionalità di archiving nella **tape library**, dove rimangono in situazione **near-line**, cioè disponibili e raggiungibili nella **tape library**, senza necessità di reperire cassette da un magazzino;
- una volta che li ha archiviati, TPI provvede a cancellare i file dal **file system** su disco;
- l'allineamento tra sito primario e sito di **Disaster recovery** viene garantito da un job periodico schedulato sul file server del sito primario, che aggiorna automaticamente il **file system** del sito secondario. Il job invia al sito secondario i nuovi file pervenuti nel **file system**, senza replicare le cancellazioni effettuate in seguito all'archiviazione su cassetta;
- sul sito di **Disaster recovery**, in maniera indipendente da quanto avviene sul sito primario, ma con politiche analoghe, viene eseguito un job di archiviazione analogo a quello del sito primario, mantenendo così l'indipendenza tra i due siti per quanto riguarda l'archiviazione.

Le funzionalità di Archiviazione di SacER verificano lo stato degli Oggetti-dati nei due siti e lo registrano sul Data Base Oracle.

Presso il sito di **Disaster recovery** viene prodotta una seconda copia per ogni cassetta. Le attività di gestione del sito di **Disaster recovery** sono tracciate in uno specifico Data Base del sistema di gestione della **tape library**.

2.1.3.7. DIPS

Il software DIPS (DIPSpenser) consiste di un sistema generalizzato in grado di configurare tramite opportuna parametrizzazione i criteri da utilizzare nella ricerca e la presentazione dei risultati quando le funzionalità di ricerca messe a disposizione dall'interfaccia web di SacER non riescono a soddisfare le particolari esigenze dell'*Utente*

Il modulo, previo controllo dei diritti di accesso alle informazioni, consente di attivare ricerche complesse sul Sistema sulla base delle **Informazioni descrittive** memorizzate dalle funzionalità di Gestione dati di Sacer e di ottenere l'*esibizione* dei documenti individuati dalla

⁵ https://developers.italia.it/it/software/r_emiro-regioneer-parer-verifica-firma-crypto.html

ricerca richiedendo gli AIP alle funzionalità di Archiviazione, sfruttando le funzionalità di Accesso di SacER.

2.1.3.8. SIAM

Il modulo software SIAM (SacER Identity and Access Management) consente di gestire varie funzionalità di amministrazione generale del Sistema, tra cui le principali sono l'autenticazione e la profilatura degli utenti (persona fisica e autonomi) e la gestione degli enti produttori e dei relativi accordi, inclusa la gestione della fatturazione.

Per le funzionalità di autenticazione SIAM utilizza sistemi di **IdP** (Identity Provider). L'Amministrazione utilizza come proprio **IdP** il sistema Keycloak ma accetta anche l'autenticazione effettuata tramite SPID (personale e professionale).

La profilatura degli utenti, sia automi che persona fisica, viene utilizzata da SacER e dagli altri moduli software del Sistema per valutare a quali viste specifiche di dati e a quali attività ogni operatore abbia accesso, sulla base dei ruoli assegnati.

SIAM mantiene il Data Base degli operatori dell'**IdP** di ParER, nonché il Data Base dei profili di tutti gli operatori abilitati al Sistema, qualunque sia l'**IdP** su cui si sono autenticati, gestendo quindi in modo centralizzato la profilatura per tutti i moduli del Sistema.

La profilatura si spinge fino al livello delle singole attività previste dal Sistema (p.e. pressione di uno specifico bottone di una specifica videata) ed al livello elementare dei dati gestiti (**Struttura**, **Unità documentaria**, Registro, ecc.) tramite la definizione e la combinazione di opportuni ruoli.

SIAM consente di gestire l'anagrafe degli enti che utilizzano il sistema di conservazione e dei loro rapporti formali (accordi) con l'ente conservatore, gestendo in particolare le fasi di avvio, esecuzione e cessazione degli accordi con i Produttori e il calcolo dei costi sostenuti sulla base dei dati acquisiti da SacER.

2.1.3.9. Sacerlog

Il modulo Sacerlog raccoglie e conserva nel sistema informazioni essenziali sul processo di conservazione in base al paradigma proposto da PREMIS, basato sui concetti di Agente (p.e. l'utente collegato al sistema), Evento (p.e. "Inserimento" o "Cancellazione") e Oggetto (p.e. "Parametro di configurazione della **Struttura**").

In pratica, ogniqualvolta un agente scatena un evento che modifica un oggetto (inclusa la creazione dell'oggetto stesso), il sistema di log registra la fotografia dell'oggetto modificato e le informazioni essenziali sulla modifica (agente, evento, timestamp, ecc.).

Il sistema di log è completamente parametrabile tramite funzioni di amministrazione, che consentono di stabilire quali combinazioni di agenti / eventi / oggetti / debbano essere registrati nel log.

Il log può essere consultato da un utente che possiede le dovute abilitazioni per determinare la storia di quanto accaduto su un oggetto. Il log può essere anche consultato per esporre il contenuto dell'oggetto ad un qualunque istante di riferimento, determinando la fotografia dell'oggetto più recente rispetto all'istante di riferimento.

2.1.3.10. Componenti di supporto per il monitoraggio

Completano il Sistema i vari moduli di supporto, ovvero le componenti che non implementano specifiche logiche applicative, ma mettono a disposizione funzionalità trasversali agli altri moduli. Più nello specifico, i principali sono:

- **Modulo di Audit e Log di sistema.**

È costituito da un insieme eterogeneo di componenti che si occupano di raccogliere tutte le informazioni rilevanti sugli eventi accaduti durante la vita del sistema. Si tratta di informazioni sistemistiche (*log di sistema* operativo, del data base e degli application server), di sicurezza (accessi andati a buon fine e rifiutati), che vengono raccolte dai diversi strati tecnologici del Sistema con il supporto di componenti specifici, ivi incluso Sacerlog.

Il modulo di Log si basa su un sistema **SIEM** (HP ArcSight) opportunamente configurato e alimentato, che si occupa di raccogliere i *log* e memorizzarli in conformità con le politiche definite da ParER sulla base della normativa vigente, con i disciplinari regionali in materia di sicurezza informatica e con la necessità di mantenere nel Sistema tutte le informazioni necessarie a documentare le attività svolte, anche per funzionalità di audit. Le informazioni memorizzate in ArcSight vengono analizzate continuamente in remoto dal Centro Operativo per la Sicurezza (**SOC**), utilizzando opportuni strumenti atti a individuare tempestivamente eventuali minacce per la sicurezza del servizio. Componenti di audit e log sono presenti anche nel sito di **Disaster recovery**;

- **Modulo di Monitoraggio Tecnico**

È costituito da un insieme eterogeneo di componenti che si occupano di raccogliere in tempo reale le segnalazioni di possibili malfunzionamenti dell'infrastruttura e di segnalarle alla struttura della Regione Emilia-Romagna preposta alla gestione.

Allo scopo vengono utilizzati diversi software open source tra cui Zabbix, che viene alimentato da opportuni agenti software, e Graylog, che raccoglie temporaneamente i log dei sistemi e li indicizza, per agevolare la diagnostica delle problematiche applicative. Le informazioni di monitoraggio sono raccolte per analisi di periodo nel cruscotto realizzato tramite un'applicazione di **business intelligence** costruita con lo strumento open source SpagoBI. Componenti di monitoraggio tecnico sono presenti anche nel sito di **Disaster recovery**;

2.1.3.11. Componenti di supporto allo sviluppo

Vengono utilizzati per garantire la corretta gestione degli sviluppi del software, per quanto riguarda sia l'evoluzione che la manutenzione correttiva del sistema.

- **Version Control System:** per l'inserimento del codice sorgente sono utilizzati GitLab per il codice usato dall'Amministrazione e GitHub per il codice pubblicato a riuso;
- **Issue Tracking System:** la tracciatura del processo di sviluppo e di tutte le attività significative connesse, dal momento della definizione dei requisiti, alla segnalazione dei bug fino al momento della richiesta di rilascio, è supportata dal software open source Redmine, su cui sono definiti diversi workflow in ragione della problematica da affrontare.

2.1.4. Sistema di gestione documentale

I servizi previsti nel presente capitolato possono essere richiesti per operare anche nell'ambito del Sistema di gestione documentale dell'Amministrazione, in particolare per lo sviluppo, la manutenzione e la gestione delle funzionalità di integrazione del Sistema di gestione documentale con il Sistema di conservazione SACER.

Il Sistema di gestione documentale in uso presso l'Amministrazione, e i relativi flussi di dati, è esemplificato in figura.

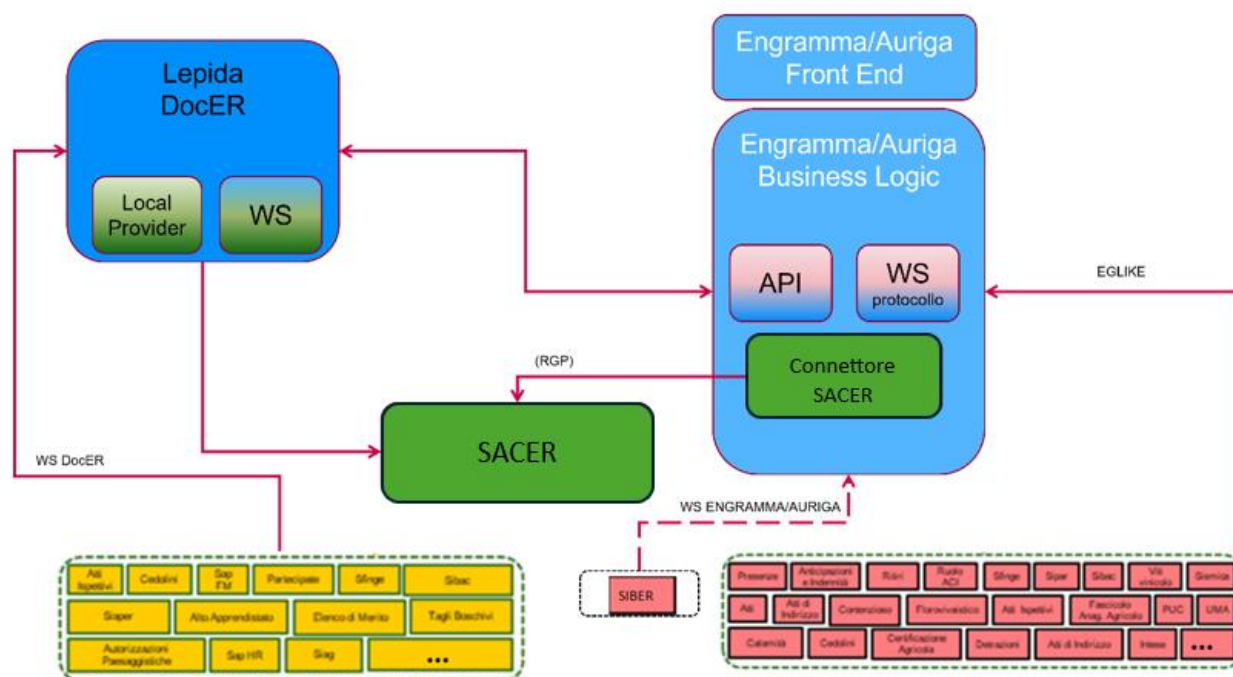


Figura 4 - Flussi di dati tra il sistema di gestione informatica dei documenti e il sistema di conservazione

Il Sistema è composto da diversi applicativi verticali produttori di documenti che, attraverso l'integrazione di un middleware di servizi documentali con il Protocollo informatico, concorrono alla formazione dell'archivio corrente dell'Amministrazione e, in seguito al versamento nel sistema di conservazione SACER, all'archivio di deposito.

Il Sistema dispone di un repository centralizzato gestito dal middleware DocER (prodotto da Lepida Scpa e installato nei sistemi dell'Amministrazione) che espone dei servizi documentali che possono essere invocati dai singoli verticali per protocollare, classificare, fascicolare e memorizzare i documenti e le aggregazioni prodotte.

A sua volta il Sistema di protocollo informatico Engramma/Auriga (sviluppato dall'Amministrazione e installato sui suoi sistemi) espone sia servizi documentali conformi alle specifiche di integrazione con DocER analoghi a quelli esposti da DocER per l'integrazione dei verticali, sia servizi documentali nativi.

I primi vengono invocati da DocER per soddisfare le richieste che pervengono dai verticali; i secondi sono invocati direttamente dai verticali.

In figura, i diversi verticali sono suddivisi a seconda del metodo di integrazione con il Sistema di gestione documentale:

- in giallo i verticali che usano i servizi nativi del sistema di protocollo Engramma;
- in rosso i verticali che usano i servizi del middleware DocER;
- i verticali che usano i servizi del sistema di protocollo e i servizi DocER.

I documenti e le aggregazioni, corredati dei relativi metadati di registrazione, descrittivi, gestionali e archivistici, confluiscono in questo modo nel repository centralizzato di DocER e sono versati a Sacer utilizzando un modulo applicativo specifico. Una minima parte dei documenti viene versata a Sacer, direttamente da Engramma/Auriga.

2.1.5. Sistema Sestra

SESTRA è una banca dati realizzata per la consultazione degli archivi storici giudiziari relativi ai processi per strage e terrorismo e della Uno Bianca di competenza della Corte D'Assise di Bologna.

È una applicazione web che dispone di funzionalità di ricerca full-text e di interfacce di consultazione dei documenti digitalizzati in PDF e sottoposti a procedure di OCR (Optical Character Recognition), inoltre, offre strumenti per l'organizzazione e la descrizione dell'archivio e per la gestione dei profili di accesso al sistema.

Il Sistema può essere acceduto dall'esterno sono dalle postazioni di lavoro presenti negli Archivi di Stato e di altre strutture pubbliche espressamente autorizzate dal Ministero della Cultura. Per questo motivo è prevista l'autenticazione tramite certificato SSL del client di accesso, oltre all'autenticazione degli utenti autorizzati, sempre dal Ministero della Cultura, ad accedere alla banca dati.

L'applicazione fornisce una serie di funzionalità per gestire un complesso documentale di provenienza analogica digitalizzato:

- Acquisizione dei documenti PDF e del testo risultante dalle procedure OCR
- Gestione delle pagine PDF
- Organizzazione delle singole pagine PDF in strutture ad albero
- Interfaccia utente per la ricerca full-text e la ricerca avanzata sui contenuti delle immagini PDF acquisite
- Interfaccia di consultazione delle unità documentali, in formato PDF e plain-text
- Gestione delle schede di descrizione archivistica

Il flusso di lavoro è esemplificato in figura.

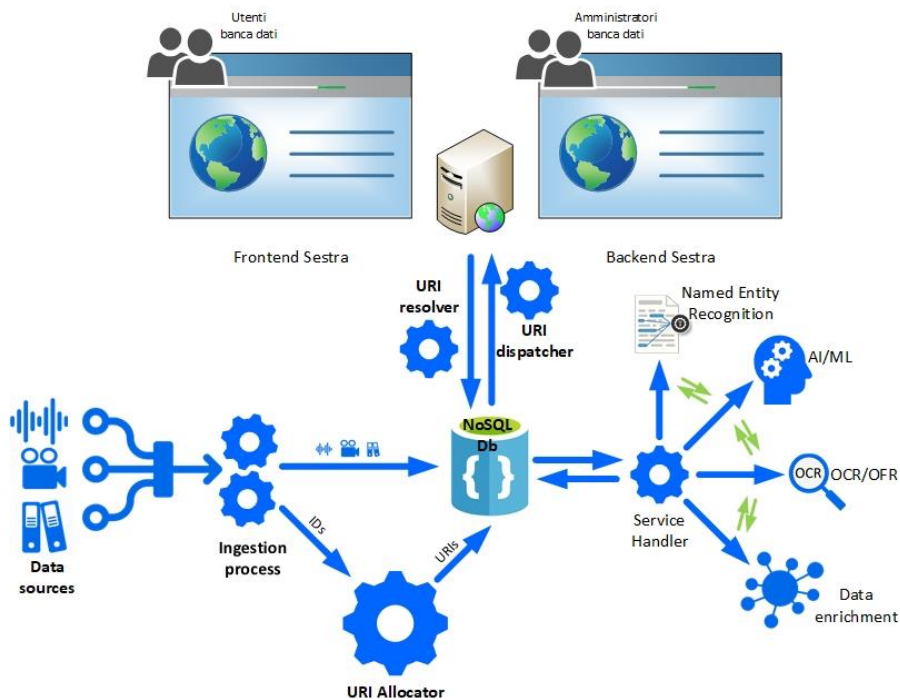


Figura 5 - Flusso dei dati del sistema Sestra

Il punto di partenza del processo è la presa in carico dei fascicoli processuali secondo un procedimento di volta in volta definito dall'Amministrazione di concerto con gli altri soggetti coinvolti e è assistito da specifiche funzionalità di metadattazione messe a disposizione dell'utente dal frontend dell'applicazione.

Nella fase di ingestion (ingestion pipeline) i singoli oggetti digitali vengono identificati con il supporto di librerie open source (ad esempio, Apache Tika), viene definito un set di metadati per ciascun oggetto digitale e in base agli stessi generare ID dei file che verranno utilizzati nell'ambito della grammatica di composizione degli URI persistenti sui quali si basa la struttura dati di Sestra.

Lo stack tecnologico è composto da: MongoDB, Elasticsearch, Apache Tomcat e applicazioni java (Storage-server, docway5, acl, docway5ws).

Il formato logico e fisico dei documenti è JSON.

Le informazioni strutturate e non sono gestite in MONGODB e l'accesso a tali informazioni e ai relativi è gestito dal motore di full-text retrieval ELASTICSEARCH, in grado di combinare ricerche istantanee sia sui dati strutturati che testuali.

3. Contesto Tecnologico

La Piattaforma è progettata e realizzata in maniera da fornire un'elevata continuità di servizio, garantire l'integrità degli oggetti conservati, gestire grandi volumi di dati, mantenere performance stabili indipendentemente dai volumi di attività e assicurare la riservatezza degli accessi.

È sviluppata con tecnologie di larga diffusione open source o comunque di libero utilizzo, a parte i sistemi di memorizzazione di dati, per i quali si utilizzano prodotti proprietari, che dispongono però di interfacce standard de facto o de jure; in particolare il Data Base per ragioni di sicurezza e di performance è proprietario (Oracle), ma standard SQL, l'Object Storage (NetApp) è proprietario, ma adotta le specifiche pubbliche S3, mentre il sistema di gestione dello storage su cassetta (TSM) è fornito da IBM, fornitore della tape library.

Il principale linguaggio di sviluppo utilizzato è **Java** su sistemi operativi Linux (Red Hat) utilizzando i seguenti componenti principali:

- Bilanciatore di carico LBL (Oplon), che svolge anche il ruolo di Web server;
- File Server GlusterFS per la gestione del file system condiviso tra diversi server e delle aree FTP;
- FTP server in cluster;
- Application server JBoss Enterprise (Red Hat) in cluster logico gestito dai componenti di clustering di JBoss;
- Servlet container Tomcat (Apache) per i componenti che non richiedono l'utilizzo di un application server (p.e. TPI, DPI, Kettle server);
- Open Shift (Red Hat) come orchestratore dei container in cui vengono eseguiti i Microservizi;
- Data Base Oracle con utilizzo delle funzionalità di RAC, di Data Guard e di partitioning;
- Object Storage NetApp distribuito in replica su tre siti;
- Tivoli Storage Manager (IBM) con funzionalità di Backup e Archiving su tape library.

I moduli applicativi della Piattaforma, essendo sviluppati in Java secondo le specifiche Java Platform Enterprise Edition (J2EE), sono raggruppati in diversi contesti applicativi caricati su JBoss.

I moduli principali sono ognuno connesso ad un proprio schema di Data Base, in modo da garantire una buona modularità dell'applicativo. I componenti che non dispongono di proprio schema di Data Base utilizzano gli altri schemi, accedendo tramite Web Service appositamente ottimizzati per l'accesso ai dati, oppure tramite opportuni grant.

I componenti applicativi che richiedono un'esecuzione fortemente dinamica sono invece sviluppati tramite Microservizi caricati in opportuni contenitori la cui gestione viene orchestrata da Open Shift.

Il colloquio tra la Piattaforma e gli applicativi esterni è effettuato tramite Web Service. Il trasferimento dei dati sincrono è realizzato in HTTPS tramite tecnologie ReST, mentre il trasferimento asincrono utilizza tecnologie FTPS.

La Piattaforma è Web-based e testato per diversi browser (Firefox, Explorer, Chrome). Non richiede l'installazione di alcun componente sul client.

Il framework di sviluppo utilizzato è stato derivato dal framework open source Spring, migliorandone gli aspetti di accessibilità; la persistenza è gestita tramite EJB, generati con la piattaforma Hibernate e solo in pochissimi casi particolari ben identificati e documentati tramite chiamate dirette JDBC, in modo da garantire portabilità verso altri Data Base relazionali e quindi facilitare il riuso dell'applicativo.

La Piattaforma ingloba diverse librerie applicative open source, molte delle quali sviluppate nell'ambito di progetti internazionali, in particolare per la verifica delle firme e dei formati.

3.1. Infrastruttura

Il diagramma in figura schematizza dal punto di vista tecnologico le principali componenti della Piattaforma e le principali relazioni con altri sistemi interessati dal processo di conservazione.

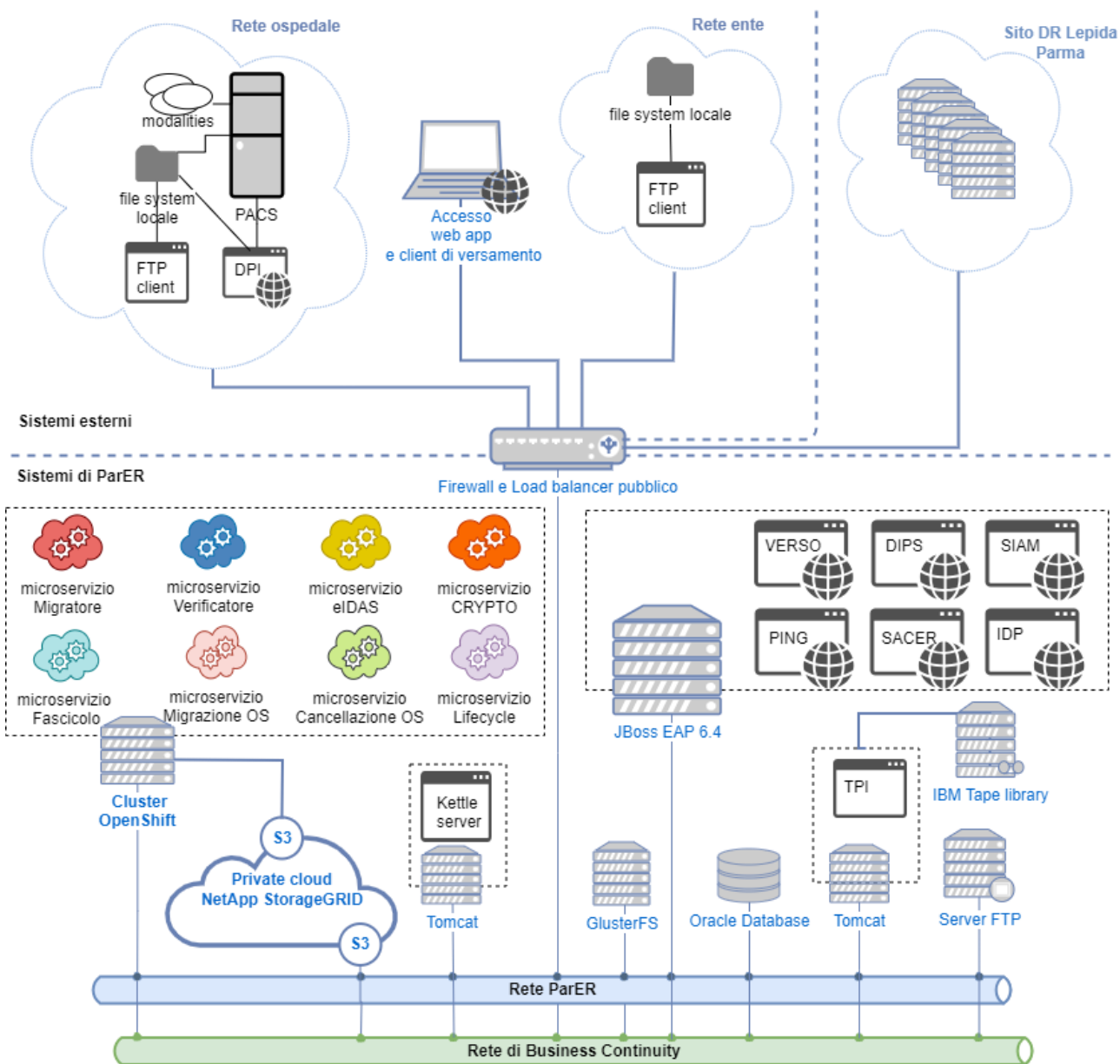


Figura 6 Architettura Infrastrutturale

3.2. Linguaggi e Tecnologie

Linguaggi di programmazione usati:

Framework	Linguaggio
JavaEE	Java
Spagolite, Spring, Spring Boot, Spring Security, Struts	Java
JPA (Hibernate, Eclipselink)	Java

JMS	Java
JSP con taglib	Java
Oracle	SQL

Altre tecnologie utilizzate:

- Bootstrap
- NetApp
- MinIO
- JBoss EAP
- Tomcat
- Openshift
- Apache NetBeans
- Docker
- Pentaho Data Integration (aka Kettle)
- Red Hat Quarkus
- JWT

3.3. Framework Spagolite

Il framework Spagolite è costituito da un set di moduli (e relative dipendenze) per "standardizzare" e "personalizzare" sviluppi legati ad applicazioni con front-end basate sul modello MVC (Model View Control) che eredita design pattern già noti da framework esistenti come Struts e Spring.

Lo scopo è quello di definire il framework con il quale le applicazioni web che prevedono front-end (e.g. Sacer), possono definire le modalità di interazione client-server, ereditare il set di librerie standard previsto dalla logica, le risorse statiche previste per la definizione di stili, javascript, immagini e molto altro, attraverso delle implementazioni standard con le quali si definisce l'intero processo.

Le applicazioni necessitano di questo framework per la loro manutenzione correttiva e per le evolutive richieste su questi moduli.