PROCEDURA PER IL MONITORAGGIO DEL PROGRAMMA OPERATIVO NAZIONALE INIZIATIVA OCCUPAZIONE GIOVANI

ALLEGATO 2

SICUREZZA E QUALITÀ DATI DELLA BANCA DATI ANPAL_PRD

SOMMARIO

Introduzione

La Base dati delle Politiche Attive del Lavoro (PROD_PRD), raccoglie i dati relativi ai giovani che hanno aderito al piano Garanzia Giovani (tramite la Scheda Anagrafica Professionale), il loro stato all'interno del progetto Garanzia Giovani (tramite le informazioni sull'Adesione, Stato Adesione), così come i dati relativi alle politiche, servizi e misure ad essi erogate (sempre tramite la Scheda Anagrafica Professionale, Flusso CO incentivate, etc.).

L'attuale scenario infrastrutturale vede il parco applicativo di ANPAL esercito e gestito dal punto di vista di tutti gli aspetti sistemistici dal fornitore del contratto SPC Lotto 1.

ANPAL ha completato le attività di reingegnerizzazione del parco applicativo e di migrazione dati al fine di dotarsi di un proprio sistema indipendente dal MLPS.

ANPAL, in aderenza al piano triennale AGID per l'informatica nella Pubblica Amministrazione, ha aderito ai contratti quadro SPC Cloud Lotto 3,4 e 1. I Lotti 3 e 4 per la realizzazione della piattaforma di interoperabilità e di quella applicativa ed il Lotto 1 per l'implementazione della piattaforma infrastrutturale ed esercizio delle applicazioni.

1. Flussi di alimentazione

La Banca Dati ANPAL_PRD viene alimentata dai SIL Regionali, tramite cooperazione applicativa o tramite accesso puntuale dell'operatore, che inviano le informazioni al Nodo di Coordinamento Nazionale, che si occupa di validarli e centralizzarli: la costante e corretta alimentazione delle banche dati da parte dei SIL regionali è fondamentale al fine di garantire un corretto monitoraggio.

2. Standard

La correttezza e coerenza sintattica dei dati inviati al Nodo di Coordinamento Nazionale viene garantita col rispetto delle *Regole di Trasmissione* alle quali sono obbligati ad attenersi tutti i nodi periferici durante l'invio dei dati al Nodo di Coordinamento nazionale.

Tali regole sono formalizzate in *standard* tecnologici e comprendono: Accordi di Servizio, Standard Tecnici e Modalità tecniche tra il nodo centrale e i nodi periferici (cfr. Documentazione allegata relativa agli standard tecnici). Ciò formalizza il rispetto di standard determinati in relazione ai modelli ed alla struttura prevista per i vari moduli di informazione (SAP, Adesione, Stato Adesione, etc).

3. Semantica dei dati

Per quanto riguarda i valori ammessi per le varie informazioni all'interno dei moduli, sono stati definiti e condivisi delle *Classificazioni Standard*: per ogni parametro all'interno della SAP, dell'Adesione, etc., sono ammessi solo codici riconducibili alle classificazioni standard.

4. Validazione flussi e contenuto informativo

Per ogni tipo di comunicazione/modello dati al livello del Nodo di Coordinamento Nazionale sono stati implementati una serie di controlli e verifiche automatiche che permettono di stabilire se l'informazione inviata dal nodo periferico è coerente con gli standard e le regole, altrimenti viene rigettata e viene inviata alla Regione (o comunque al nodo periferico) mittente una notifica con la segnalazione del motivo del rigetto/scarto.

Le informazioni raccolte all'interno della Banca Dati ANPAL_PRD attraverso l'esecuzione giornaliera di flussi e procedure automatizzate vengono archiviate e gestite all'interno di una Data Platform (che al livello più alto alimenta direttamente il Sistema di Monitoraggio di Garanzia Giovani).

All'interno della Data Platform le informazioni della Banca Dati ANPAL_PRD vengono aggiornate, integrate, elaborate, normalizzate e consolidate al fine di ottenere un sistema che garantisca accuratezza, pertinenza, rilevanza, trasparenza e tempestività dei dati.

In particolare, la Data Platform:

- Contiene dati coerenti e "ripuliti" da eventuali errori nei dati alla fonte, preservando la qualità stessa dei dati utilizzati per l'analisi; è garantita la rispondenza dei dati alle norme e agli Standard definiti;
- I dati esposti possono essere ricombinati e separati rispetto alle esigenze di analisi, è possibilità di produrre indicatori a differenti livelli di aggregazione (che, poi il sistema di raccolta dei dati è decentrato, coinvolgendo tutte le regioni, ha implicato l'omogeneità delle classificazioni adottate);
- E' possibile produrre indicatori per scopi differenti (individuazione delle criticità e delle anomalie, monitoraggio performance di sistema);
- Viene garantita la sicurezza dei dati: è possibile consultare i dati tramite la Dashboard di Monitoraggio alla quale si può accedere tramite utenze e password di dominio. La Data Platform è un sistema in sola lettura, gli utenti non eseguono azioni di aggiunta, modifica o eliminazione dei dati.
- L'accesso ai dati è garantito con bassi tempi di attesa fra l'interrogazione dei dati e l'output di risultati.

Nella Data Platform i dati sono presenti al massimo dettaglio e storicizzati raggiungendo dimensioni poco compatibili con l'interrogazione diretta attraverso i diversi tool di analisi a disposizione: sono stati perciò generati aggregati di dati (datamart) ai quali accedono direttamente le Dashboard di Monitoraggio. Per rispondere a richieste di date di dettaglio o puntuali.

È all'interno della Data Platform quindi che vengono effettuati una gran parte di controlli semantici sulla coerenza, 'pulizia' e affidabilità dei dati ricevuti che non possono essere effettuati automaticamente a livello di nodo nazionale:

- sia perché possono emergente dal confronto tra le stesse informazioni proveniente da diverse fonti;
- sia perché possono emergere dal confronto tra i dati diversi ricevuti in modo asincrono e quindi le verifiche possono essere effettuate in un momento successivo alla ricezione;

Ad esempio a livello di NCN ci sono controlli che non permettono di ricevere adesioni relativa a soggetti con età non in fascia Garanzia Giovani al momento dell'adesione stessa, o inserire all'interno della sezione 6 della SAP una politica attiva non prevista; ma è e livello di Data Platform che viene verificata la coerenza tra: le informazioni associate al Giovane all'interno della SAP; le informazioni sullo stato Adesione del Giovane; le informazioni sull'eventuale Presa in Carico del Giovane (e quindi del calcolo del profiling).

Gli indicatori definiti per il Sistema di Monitoraggio utilizzano solo dati corrispondenti ad informazioni coerenti con il flusso previsto e la normativa (ad esempio tra i Presi in Carico vengono conteggiati solo i giovani per i quali è presente la politica A02 all'interno della Sezione 6 della SAP e per i quali è stato effettuato il calcolo del Profiling coerentemente con la A02); gli altri dati vengono gestiti come scarti ed anomalie da verificare ed eventualmente sanare con l'intervento delle Regioni: la Data Platform è infatti un sistema in sola lettura, gli utenti non eseguono azioni di aggiunta, modifica o eliminazione dei dati.

Il processo di data quality effettuata all'interno della Data Platform può mettere in luce problematiche nascoste all'interno dei processi aziendali, tipicamente l'assenza di dati relativamente ad una particolare attività dovuta all'opzionalità di compilazione dei dati (causa principale dell'assenza di dati) può evidenziare, proprio grazie alla Data Platform, la necessità di verificare a livello di processo direttamente con le Regioni la possibilità/necessità di rendere obbligatorie informazioni prima opzionali piuttosto che gestire in modo diverso alcune informazioni (es. Addendum al vademecum).

5. Procedure per il backup/recupero dati

I database interessati dal programma sono sottoposti a backup in modo da assicurare la loro ricostruzione in caso di incidente con perdita dei dati.

Gli aspetti fondamentali nella definizione delle policy di backup sono i seguenti:

- Periodicità dei salvataggi

  Il backup del database ANPAL_PRD viene effettuato ogni 24 ore.

  Unitamente ad esso viene effettuato il backup dei wal segment procedendo a lotti. Dimensione e frequenza dei lotti vengono definiti dall' rdbms e decisa dal motore del database.

- Procedure di restore

  Le procedure di restore prevedono l'impiego del sistema centralizzato di backup che può effettuare il ripristino del DB all'occorrenza

- Supporti utilizzati per il salvataggio

  I salvataggi sono conservati su apparati differenti da quelli che contengono la base dati originale in modo che la perdita di quest'ultimo intero apparato non comprometta la disponibilità delle copie di backup.

  l dato di backup è replicato verso un sito di Disaster Recovery.

- Rintracciabilità dei backup

  L'Amministratore del Sistema identifica i backup con una nomenclatura univoca e non ambigua in modo tale da poter riconoscere facilmente i singoli backup.

6. Gestione della sicurezza

La gestione della sicurezza viene suddivisa in:

- sicurezza fisica
- sicurezza logica

Per quanto riguarda la sicurezza fisica, i sistemi da cui vengono erogati i servizi sono quelli messi a disposizione dal Lotto 1 SPC Cloud. Le modalità di gestione ed accesso ai sistemi fisici sono definite e regolamentate nelle clausole contrattuali comuni a tutte le pubbliche amministrazioni aderenti.

In relazione alla sicurezza logica gli aspetti basilari sono:

  - l'utilizzo di strumenti, quali i firewall e similari, che consentano la protezione dei sistemi da accessi fraudolenti. L'attuale configurazione prevede l'utilizzo di firewall virtuali
- il cluster dove risiede il DB è su un server linux Ubuntu che è soggetto agli aggiornamenti delle patch di sicurezza
- con cadenza mensile vengono installati gli aggiornamenti di sicurezza.

7. Amministrazione e gestione delle utenze

Il processo di gestione delle utenze prevede che gli utenti applicativi abilitati ad accedere al DB, hanno il livello di accesso attribuito tramite permessi puntuali assegnati dietro richiesta effettuata per mezzo di piattaforma di Service Management che provvede a memorizzare e storicizzare l'iter di richiesta.

La gestione delle utenze interne all'Amministrazione avviene tramite i tool Microsoft e quindi si utilizza come repository l'active directory.

Per quanto riguarda le utenze esterne al momento non sono utilizzati strumenti di access management ma la verifica della correttezza dei dati utilizzati per l'accesso è demandata alle singole applicazioni.

Sono state completate le attività di reingegnerizzazione del software e della realizzazione della piattaforma infrastrutturale necessaria all'esercizio applicativo attraverso l'adesione ai contratti quadro SPC Cloud.

L'esercizio delle applicazioni e la gestione dei sistemi avverranno attraverso il Cloud SPC erogato dal RTI aggiudicatario della gara SPC Cloud Lotto 1.

In tale scenario si fa riferimento alla documentazione ufficiale SPC Cloud Lotto 1 pubblicata sul sito CONSIP all'indirizzo (http://www.consip.it/media/news-e-comunicati/gara-spc-cloud-disponibile-la-documentazione).

7.1. Procedure per il backup/recupero dati

Nell'ambito del "Servizio L1.S1.2 – Virtual Data Center" viene già effettuato di default dal fornitore di infrastruttura il backup settimanale dei volumi di boot al fine di proteggere le stesse da eventi avversi. La soluzione permette il ripristino delle VM, su richiesta di ANPAL. Viene inoltre garantita una *retention* dei backup pari a 21 giorni solari.

Più specificatamente ANPAL ha anche attivato il servizio BaaS (Backup as a Service) che consente una maggiore flessibilità ed autonomia nella definizione degli oggetti di backup e delle relative policy.

Attraverso tale servizio vengono backuppate le directory contenenti il software applicativo con una retention di 30 giorni.

Per approfondimenti si rimanda alla documentazione ufficiale SPC Cloud Lotto 1 pubblicata sul sito CONSIP all'indirizzo (http://www.consip.it/sites/consip.it/files/Gara-Cloud-Lotto-1_Allegato-A_Capitolato-Tecnico.pdf).

ANPAL ha anche attivato il "Servizio L1.S2 – Platform as a Service (PaaS)" che prevede l'erogazione di servizi middleware per lo sviluppo, collaudo, manutenzione ed esercizio di applicazioni, su infrastruttura hardware sottostante di tipo IaaS, del tutto trasparente all'Amministrazione. In particolare, il Database relazionale per i dati transazionali, erogato in modalità PaaS, scelto da ANPAL è l'RDBMS PostgreSQL presente nella Lista dei "Solution Stack". Il backup del RDBMS viene effettuato dal fornitore infrastrutturale attraverso il software Barman completamente gestito dal fornitore infrastrutturale (https://www.pgbarman.org)

7.2. Gestione della Sicurezza

Relativamente alla Sicurezza fisica, i Centri Servizi, e in generale tutte le sedi aziendali, dispongono di misure avanzate di protezione attiva e passiva (recinzioni, sbarre, cancelli, dissuasori, dispositivi di controllo accessi) nonché organizzativa (vigilanza, reception, ecc.).

I Centri Servizi hanno livelli multipli di protezione, attrezzati con meccanismi quali:

- controllo accessi, differenziato per aree sensibili quali le sale sistemi, abilitato da meccanismi tradizionali (es. badge) e meccanismi più avanzati (riconoscimento biometrico);
- sistemi anti-intrusione, quali recinzioni, vigilanza H24, telecamere a circuito chiuso;
- rack (armadi che ospitano i sistemi) protetti da gabbie fisiche con serrature a chiave;
- impianti di continuità elettrica, rilevazione fumi, spegnimento incendi e antiallagamento tali da proteggere i sistemi.

Relativamente alla Sicurezza logica viene garantita ad ANPAL l'isolamento e la protezione dei dati. I sistemi operativi sono conformi alle specifiche indicate nella circolare AIPA n. 31 del 21/6/2001 e successive modificazioni e le registrazioni di sicurezza sono protette da modifiche non autorizzate (DPCM 31/10/2000, Art. 7, comma 4 e successive modificazioni). Le tematiche di gestione del rischio e della compliance vengono indirizzate attraverso:

- processi di Gestione della Sicurezza che saranno descritti nel dettaglio nel Piano della Sicurezza;
- corretta gestione dei profili di accesso di tipo amministrativo e separazione organizzativa tra le funzioni deputate all'assegnazione di credenziali e diritti di accesso e quelle di gestione tecnica dei sistemi, in conformità al Provvedimento del Garante Privacy 1.6.2006;
- rispetto degli obblighi previsti dal Testo Unico in materia di privacy – D.Lgs. 196/03;
- conformità agli standard internazionali di sicurezza e alle best practice richiamate anche dalla norma ISO27001 in materia di User Access Management;

Per approfondimenti si rimanda alla documentazione ufficiale SPC Cloud Lotto 1 pubblicata sul sito CONSIP all'indirizzo (http://www.consip.it/sites/consip.it/files/Gara-Cloud-Lotto-1_Allegato-B_Offerta-Tecnica-del-Fornitore.pdf).

7.3. Amministrazione e gestione delle utenze

Le utenze per l'accesso alla piattaforma infrastrutturale sono gestite dal RTI SPC Cloud Lotto 1 nell'ambito delle proprie procedure di sicurezza. Le utenze che accedono alla Piattaforma e che gestiscono i servizi 'Managed' sono classificati come Amministratori di Sistema e quindi sono stati censiti nell' apposito elenco e hanno ricevuto la lettera di incarico.

Relativamente ai database applicativi questi vengono creati, su richiesta del fornitore applicativo, dal fornitore infrastrutturale attraverso un utente super-user gestito dagli Amministratori di Sistema del fornitore. Il fornitore comunica quindi l'utenza owner dello schema all'incaricato ANPAL indicato nel Piano e Progetto dei Fabbisogni. Il fornitore infrastrutturale non è responsabile dell'utilizzo delle utenze e degli schemi creati.

Le pipeline di CI/CD implementate consentono degli automatismi di replacement delle password applicative sui file di configurazione senza che i team di sviluppatori conoscano tali credenziali di accesso.