

## **<Nevis per Irpet>**

**<proposta per “Servizi di assistenza tecnica e di consulenza statistico – informatica alla gestione e manutenzione di sistemi informativi, a supporto delle ricerche previste nel Programma istituzionale dell’IRPET per l’anno 2017 ” >**

**Vostro riferimento: Prot. n. 1811/2017 / CIG: Z991FEED88**

## **Offerta Tecnica / 21 settembre 2017**

# Indice

[Indice](#)

[Versione del documento](#)

[Scenario ed obiettivi](#)

[Attività e funzionalità da sviluppare](#)

[A1 - adeguamento della procedura di ETL da dump del sistema informativo lavoro CO/IDOL](#)

[A2 - Refresh cubi con cadenza trimestrale.](#)

[B1 - Realizzazione di una anagrafica delle unità locali di tipo longitudinale](#)

[B2 - Realizzazione anagrafica per le imprese del settore agricolo](#)

[Attività comuni ai due progetti](#)

[Adeguamento wiki di progetto](#)

[Attività sistemistiche](#)

[Team Nevis](#)

## Versione del documento

Versione	Data	Note
v. 1	21 settembre 2017	Offerta tecnica

## Scenario ed obiettivi

Il presente documento ha come obiettivo quello di fotografare le necessità di IRPET in termini di supporto statistico-informatico a sostegno dell'attività di ricerca istituzionale sui temi:

- A. mercato del lavoro (valorizzando basi di dati amministrative IDOL/COB);
- B. osservatorio imprese (dinamica e funzionamento del sistema produttivo regionale).

Relativamente al tema A l'attività prevede l'aggiornamento del database statistico inerente i flussi delle comunicazioni obbligatorie ai Centri per l'impiego, con particolare riferimento a:

- A1- aggiornamento della procedura di ETL dei dati da dump del sistema informativo lavoro CO/IDOL;
- A2- 4 refresh completi strutture analitiche OLAP/cubi (cadenza trimestrale).

Relativamente al tema B l'attività prevede la realizzazione delle seguenti due basi di dati analitiche:

- B1- anagrafica longitudinale delle unità locali di imprese Toscane (periodo 2007-2014)
- B2- anagrafica sperimentale per le imprese del settore agricolo (sezionale, un solo anno)

## Attività e funzionalità da sviluppare

Di seguito una descrizione degli interventi necessari sulla base degli obiettivi appena evidenziati a supporto delle attività di analisi per i temi A e B.

### **A1 - adeguamento della procedura di ETL da dump del sistema informativo lavoro CO/IDOL**

Una serie di interventi normativi hanno avuto impatto sui meccanismi di raccolta e gestione amministrative delle comunicazioni relative alla mobilità e alla disoccupazione; procederemo, quindi, con l'analisi dei passi necessari ad adeguare le procedure. Obiettivo ultimo è quello di garantire nuovamente adeguata copertura in termini di profondità di analisi per i fenomeni legati a disoccupazione e mobilità.

Ad ogni iterazione opereremo comunque le necessarie attività di fine-tuning delle procedure di caricamento delle basi dati sorgenti (con contestuale verifica della loro consistenza/integrità).

## **A2 - Refresh cubi con cadenza trimestrale.**

Supporteremo IRPET ad ogni iterazione per 4 scarichi completi. Il supporto consta di attività statistico/informatica utile al refresh dei dati, delle strutture analitiche e dei cubi per l'analisi dei principali aggregati. Oltre alla mera riproduzione delle strutture affiancheremo IRPET anche per le attività di lettura e validazione dei dati.

## **B1 - Realizzazione di una anagrafica delle unità locali di tipo longitudinale**

Completata la fase di costruzione della base dati longitudinale delle imprese ora è necessario affiancarle una struttura simile con *scope* le unità locali. Per permettere analisi di natura longitudinale e di coorte è necessario costruire un apposito archivio che sappia identificare le varie immagini giuridico/amministrative delle imprese (e delle unità locali); i record di anagrafica sono infatti risultato delle attività gestite dagli archivi amministrativi (e statistici) nella disponibilità del progetto e secondo le regole che alimentano i sistemi stessi. Naturale conclusione è che spesso le imprese mutano alcune caratteristiche che le portano fuori dall'universo osservato da un archivio o di contro appaiono come nuove in altri archivi per ragioni analoghe. In questa fase ci occuperemo di definire le regole di linkage più adatte a seguire i soggetti impresa nel loro manifestarsi all'interno degli archivi. Useremo tecniche di link deterministico (campi identificativi comuni agli archivi quali la PIVA o CF), probabilistico (statistical matching) e tecniche miste (fuzzy logic matching). Obiettivo ultimo è avere la possibilità di osservare il comportamento economico delle imprese e delle unità locali lungo un periodo ampio e al netto dei problemi di identificazione delle stesse.

## **B2 - Realizzazione anagrafica per le imprese del settore agricolo**

Al momento questo asse progettuale può contare sulle seguenti basi di dati:

- archivio statistico delle imprese attive (ASIA) per il settore agricoltura;
- censimento dell'Agricoltura 2010;
- dichiarazioni a fini fiscali;
- base dati Artea sulle imprese agricole che hanno ricevuto finanziamenti FEASR;
- esportazioni;
- registro delle imprese;
- indagine sulla struttura e sulle produzioni delle aziende agricole;
- indagine sui risultati economici delle aziende;
- sistema informativo lavoro.

Resta invece da verificare la disponibilità di dati Inps e Inail.

Dopo una fase preliminare di censimento delle fonti disponibili/accessibili passeremo al

loro caricamento in apposita area di staging per una valutazione della loro qualità. Le stesse verranno rese omogenee e trattate con l'obiettivo di definire regole di priorità tra le fonti durante la realizzazione della anagrafica. Una prima versione dell'anagrafica sulle imprese agricole avrà finalità prevalentemente sperimentali e servirà per dimensionare effort futuri e fattibilità di un progetto a più lungo termine.

## **Attività comuni ai due progetti**

### **Adeguamento wiki di progetto**

Tutte le fasi di analisi e sviluppo verranno accompagnate da adeguata attività di documentazione sull'ambiente (wiki) di *knowledge management* del progetto.

### **Attività sistemistiche**

Ad ogni iterazione storicizzeremo le precedenti versioni della base dati su strutture *archived* per ottimizzare l'occupazione dello spazio. Saranno sempre e comunque interamente accessibili/interrogabili, senza perdita di prestazioni, in modalità di sola lettura. Le strutture aggregate (cubi OLAP) saranno mantenute online e gestite su RDBMS colonnare ad alte prestazione HP Vertica nella versione Community Edition.

## **Team Nevis**

Il team di lavoro è composto da Danilo Maurizio e Gianluca Emireni entrambi senior consultant di Nevis.