

CORSO INTRODUTTIVO ALLA GESTIONE DEI DATI PER I PARCHI NAZIONALI ALPINI ITALIANI

MODULO 1

Introduzione alla gestione dei dati e ai database



Lezione 1. Introduzione alla gestione dei dati

OBIETTIVO DEL CORSO

Rendere i Parchi autonomi nell'uso e aggiornamento del proprio database secondo i principi della corretta gestione dei dati

Coordinatori scientifici: saper gestire lo sviluppo del DB e coordinarlo con le altre attività del Parco.

Tecnici: saper utilizzare i dati inseriti nel database.

Collaboratori: saper ottimizzare la registrazione dei dati e gestire la fase di digitalizzazione dei dati per l'inserimento nel database.

Esperti informatici: sapere cosa è necessario per inserire il database nel sistema informativo del Parco.

Curatori dei dati: saper controllare la qualità dei dati e caricamento nel DB. Saper estendere la struttura e dare supporto tecnico agli utenti.

— — —

OBIETTIVI PER OGNI TIPO DI OPERATORE

STRUTTURA DEL CORSO

MODULO 1

— — —

Introduzione alla gestione dei dati e DB relazionali

Introduce nozioni di base di gestione dei dati, spiega cos'è e come è fatto un database relazionale, mostra come collegarsi ai database, a vedere una tabella e scaricare i dati in formato csv.

MODULO 2

Introduzione al linguaggio SQL di base

Spiega come interrogare i dati in un DB tramite comandi SQL di base, come modificare i dati nelle tabelle e come scaricare i dati formattandoli a piacere.

MODULO 3

Comandi SQL per interrogazioni complesse

Spiega come interagire con i dati in un DB in modo avanzato con comandi SQL più complessi in modo da poter combinare dati di strutturati in diverse tabelle e processarli.

MODULO 4

Gestione e aggiornamento del database

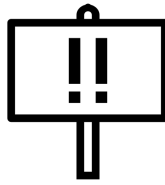
Insegna come creare nuovi oggetti nel database, processare e verificare nuovi dati raccolti sul campo e inserirli nel database dopo averli controllati e organizzati.

1. Alcuni problemi tipici nella gestione dei dati
2. Requisiti per la gestione dei dati
3. Opzioni per l'archiviazione dei dati
4. Piano per la gestione dei dati
5. Principali caratteristiche dei database relazionali
6. Perché i Parchi stanno creando il proprio database



1. Alcuni problemi tipici nella gestione dei dati

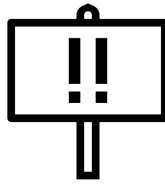
- Qualità dei dati non verificata con errori non corretti
- I dati archiviati richiedono tempo di preparazione per uso
- Dati archiviati in più versioni
- Più persone lavorano su versioni diverse
- Informazioni comprensibili solo a chi le ha raccolte
- Difficoltà a riusare i dati per altri scopi
- Dati persi sul lungo periodo



2.1 Requisiti per la gestione dei dati

— — —

- I dati devono essere completi
- I valori devono essere classi/range validi
- I valori devono essere corretti
- Le informazioni in un data set non si devono contraddire
- Ci deve essere una sola versione dei dati
- I dati/protocolli/metodi devono essere documentati
- I dati devono essere accessibili
- L'accesso ai dati deve essere protetto



2.2 Requisiti per la gestione dei dati

- Si deve poter estrarre i dati in base alle esigenze
- Più persone devono poter lavorare ai dati contemporaneamente
- I dati devono essere conservati per lunghi periodi
- I dati da gestire sono vari: tabulari, spaziali, temporali, ...
- I dati sono usati con tool diversi: stats, report, GIS, ...
- La gestione dei dati deve richiedere risorse ragionevoli
- *La gestione dei dati deve essere fatta con risorse disponibili*

Una appropriata gestione dei dati costa il 5-10% del processo di acquisizione, ma migliora molto l'impatto del loro uso/riuso.



3.1 Opzioni per l'archiviazione dei dati

- CSV file e documentazione (e.g. csv e documenti di testo)
- Fogli di calcolo (e.g. MS Excel, LibreOffice Calc)
- Database locali (e.g. MS Access, Sqlite)
- Archivi web condivisi (e.g. GBIF, Ornitho, Movebank)
- Database centralizzati (e.g. PostgreSQL, Oracle, MySQL)



3.2 Opzioni per l'inserimento dei dati in un database

— — —

Quando i dati vengono archiviati in un database centralizzato, ci sono varie opzioni per inserire i dati nel database, ad esempio:

- Tablet con applicazioni apposite collegate al DB (dati da validare)
- Adattamento di applicazioni esistenti (e.g. Ornitho, ODK)
- Tablet, scaricamento dei file, validazione e upload nel DB
- Utilizzo di interfacce web per l'inserimento dei dati dalle schede
- Fogli di calcolo strutturati per inserimento da schede e poi upload nel DB
- Fogli di calcolo non strutturati per inserimento da schede e upload nel DB
- Inserimento manuale diretto dei dati nel DB senza maschere

I protocolli di raccolta, la registrazione dei dati, la struttura dati nel database e la procedura di informatizzazione devono essere **coordinate** come fasi di uno stesso processo.



3.3 Fasi di processamento dei dati storici

Per i dati storici, tipicamente archiviati in fogli di calcolo non strutturati, le fasi di lavoro per integrarli in un database sono:

1. **Validazione** dei dati (fatta dal curatore dei dati per il controllo della qualità assieme a chi ha raccolto i dati)
2. Creazione della **struttura dati** nel database (fatta dal gestore del database)
3. L'**importazione** dei dati nel database (fatta dal curatore dei dati)
4. La **documentazione** dei dati (campi, struttura, protocollo di raccolta)



4.1 Piano per la gestione dei dati: cos'è

- Un **Data Management Plan** è un documento da preparare all'inizio di un progetto (e poi da aggiornare) che descrive come devono essere gestiti i dati durante e dopo la loro raccolta.
- Fornisce una descrizione dei dati, dei protocolli di raccolta, degli standard usati, delle politiche di accesso, della conservazione di dati, e dei responsabili delle varie attività.
- Garantisce che i dati siano nel formato corretto, ben organizzati e meglio annotati.
- I dati potranno essere compresi anche da chi non li ha raccolti e riutilizzati anche in futuro.
- I dati possono essere formattati durante la raccolta per rendere più facile l'inserimento in un database.

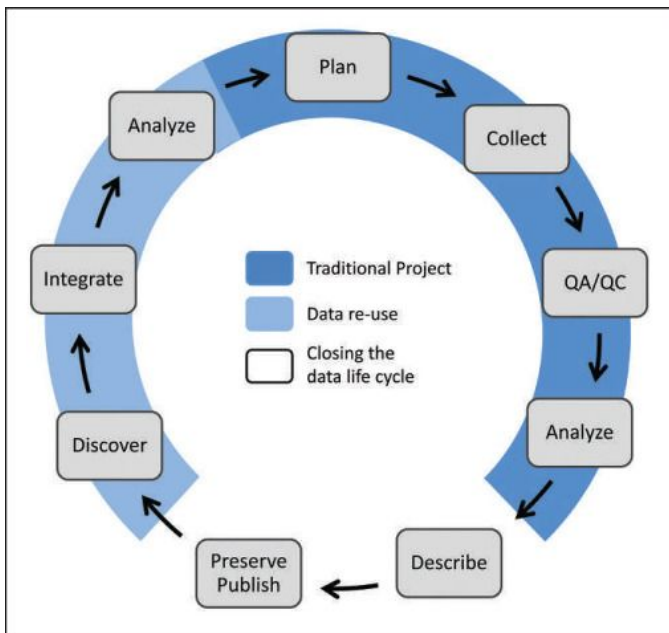


4.2 Piano per la gestione dei dati: elementi chiave

Per garantire che i dati siano di qualità, preservati e utilizzabili, la gestione dei dati deve essere una delle attività integranti delle attività dei Parchi

1. Acquisizione dei dati
2. Controllo di qualità dei dati
3. Archiviazione dei dati
4. Aggiornamento dei dati
5. Accessibilità dei dati
6. Responsabile dei dati
7. Costo per la gestione dei dati

4.3 Piano per la gestione dei dati: ciclo di vita dei dati



1. **Pianificare:** definizione dei dati che saranno raccolti e come saranno gestiti
2. **Raccogliere:** osservazioni di operatori o sensori e digitalizzazione
3. **QA/QC:** la qualità dei dati assicurata con controlli e strutturata nel formato/strumento adeguato
4. **Analizzare:** i dati vengono analizzati/utilizzati (output scientifici o gestionali)
5. **Descrivere:** i dati sono documentati utilizzando gli appropriati standard per metadati
6. **Conservare:** i dati sono presentati a un appropriato archivio a lungo termine
7. *Scoprire: i dati vengono trovati e acquisiti, insieme alle informazioni rilevanti sui dati (metadati)*
8. *Integrare: i dati provenienti da fonti diverse sono combinati per essere facilmente analizzati*
9. *Analizzare: i dati vengono analizzati/utilizzati*



5.1 Principali caratteristiche dei database relazionali

- I tipi di dato sono definiti in modo esplicito
- I valori ammessi dei dati possono essere ristretti
- Gli oggetti e le loro relazioni sono formalizzati
- Non c'è ridondanza dei dati (normalizzazione)
- Si può accedere al database da remoto
- Viene gestita la concorrenza di accessi



5.2 Principali caratteristiche dei database relazionali

- Si possono assegnare livelli di accesso diversi
- Si possono fare backup automatici
- Si possono archiviare enormi quantità di dati
- Possibilità di interrogare i dati in modo complesso
- I dati vengono processati velocemente



5.3 Principali caratteristiche dei database relazionali

- Il sistema è centralizzato (richiede un server)
- Ha una struttura server/client (piattaforma modulare)
- Si possono utilizzare i dati con qualsiasi software
- I dati sono archiviati in modo sicuro
- Standard industriali per i formati dei dati
- Interoperabilità tra diversi database



5.4 Principali caratteristiche dei database relazionali

— — —

- Tecnologia consolidata
- Gestione dei dati spaziali
- Gestione dei dati temporali
- **Necessarie conoscenze di base per utilizzo**
- **Necessarie conoscenze avanzate per gestione**



5.5 Database relazionali: sintesi

- Garantisce l'integrità dei dati
 - Si prevengono errori di inserimento
 - Si hanno solo informazioni "pulite"
 - Si formalizza l'informazione
- Aumenta i livelli di sicurezza
 - I dati non vengono corrotti per errori degli operatori
 - I dati non possono essere usati da non autorizzati
- Permette il riutilizzo dei dati
 - Uso dei dati per diverse applicazioni
 - Uso dei dati sul lungo termine
 - Integrazione dei dati con altri data sets
 - Documentazione dei dati



5.6 Database relazionali: sintesi

- Previene la duplicazione dei dati e migliora accesso
 - I dati possono essere usati da più persone insieme
 - I dati possono essere usati da dovunque
 - I dati possono essere facilmente condivisi
- Si possono gestire dataset grandi e complessi
 - Grande capienza di storage
 - Modelli di dati con relazioni formali complesse
 - Performance nell'uso dei dati

6.1 Perché i Parchi stanno creando il proprio database



- Rendere fruibili i dati raccolti (adesso e in passato)
- Preservare i dati sul lungo periodo
- Verificare e migliorare la qualità dei dati
- Armonizzare i dati raccolti all'interno di ogni Parco
- Conservare i dati in un unico archivio
- Rendere più efficiente la gestione dei dati
- Facilitare la connessione con altri progetti/network/istituzioni
- Semplificare l'eventuale integrazione con dati di altre fonti (remote sensing, modelli climatici, etc)
- **PERCHÉ DI FATTO NON CI SONO ALTERNATIVE** (sul medio e lungo periodo) compatibili con il mandato dei Parchi



6.2 Progetto Biodiversità (gestione dei dati)

I dati del Progetto Biodiversità sono stati tutti integrati nei database dei rispettivi Parchi

- La struttura dei data set nei vari database è consistente
- I dati sono stati armonizzati
- I dati hanno avuto simili processi di verifica della qualità
- La coerenza dei protocolli di raccolta dati fra Parchi è stata verificata
- Le procedure di registrazione dei dati sono state ottimizzate
- I database dei parchi sono separati ma interoperabili
- I dati sono condivisi in un database “virtuale” comune
- Esperienza pilota per future collaborazioni fra i 4 Parchi
- Esperienza pilota per future collaborazioni anche con altri Parchi

Fine Lezione 01

Prossima lezione: [Introduzione ai database](#)