

DEFINIZIONE DI BUONE PRASSI PER LA CONSERVAZIONE DEI SITI WEB DI PROGETTO O DI LABORATORIO AI FINI DELLA LORO CONSERVAZIONE AL TERMINE DELL'EVENTO CHE LI HA GENERATI

INTRODUZIONE	2
RICOGNIZIONE DELLA SITUAZIONE ATTUALE	2
IL CICLO DI VITA DI UNA WEB APPLICATION	3
<i>La nascita: attivazione.....</i>	<i>3</i>
<i>La maturità: sito pubblico, funzionante e mantenuto</i>	<i>4</i>
<i>Chiusura, reingegnerizzazione, staticizzazione o limitazione alle funzionalità</i>	<i>5</i>
CONSERVAZIONE DI SITI WEB: STRUMENTI TECNICI, STANDARD INTERNAZIONALI E BUONE PRATICHE....	6
CREAZIONE DELLA COPIA STATICA DI UN SITO	6
IL FORMATO WARC	6
SISTEMA DI GESTIONE DEI SITI ISTITUZIONALI DI CENTRI/LABORATORI E GRUPPI DI RICERCA	7
<i>Indirizzi di ricerca.....</i>	<i>8</i>
<i>Eventi.....</i>	<i>8</i>
<i>Pubblicazioni</i>	<i>9</i>
<i>Persone.....</i>	<i>9</i>
<i>Progetti.....</i>	<i>10</i>
ATTUAZIONE DELLE LINEE GUIDA A UN CASO DI STUDIO: CHIUSURA DEL CENTRO DI FILOSOFIA	10

INTRODUZIONE

Scopo di questo documento è la determinazione di un protocollo finalizzato a sviluppare i processi gestionali e metodologici per la conservazione sostenibile dei siti web – di progetto o istituzionali – della Scuola al fine di mantenere il livello più alto livello di sicurezza della rete.

In particolare con la dicitura protocollo per una “conservazione sostenibile” si intende l’insieme delle **procedure informatiche e amministrative** indirizzate a favorire la sopravvivenza dei siti web al termine dell’evento che li ha generati (per es. scadenza del progetto, conclusione del mandato del responsabile scientifico, conclusione di un laboratorio o struttura accademica) o al termine del loro naturale ciclo di vita (es. siti obsoleti, siti che presentano caratteristiche per cui è impossibile l’aggiornamento dei componenti o del sistema), preservandone gli standard di sicurezza della Scuola.

Con il presente documento, a partire dall’analisi della situazione attuale dei siti web della Scuola, ci si propone di definire le linee guida relative alla gestione dell’esistente e un protocollo su cui basarsi per i siti web di nuova attivazione, definendo così le modalità di gestione di siti e server della Scuola per l’intero *ciclo di vita dei siti web*.

E’ inoltre descritto un caso specifico in cui sono state messe in atto le **procedure informatico-amministrative** di gestione. Il caso in oggetto ha riguardato i siti del Centro di Filosofia, che ha terminato la sua attività di ricerca e che ha lasciato in eredità alla Scuola diversi siti di progetto gestiti dal personale del laboratorio stesso.

Le linee guida proposte in questo documento di analisi sono destinate alla pubblicazione e alla divulgazione attraverso il sito web dei servizi informatici della Scuola.

RICOGNIZIONE DELLA SITUAZIONE ATTUALE

Alla Scuola si evidenzia una situazione relativa alla gestione dei siti web molto variegata e distribuita:

- i siti istituzionali e una parte di quelli di progetto sono gestiti direttamente dall’Area Strategie Digitali (su server residenti in data center interni o esterni);
- i siti di ricerca o relativi a progetti finanziati sono generalmente gestiti in autonomia da laboratori, centri e gruppi di ricerca su server propri o su data center esterni;
- siti dei collegi sono gestiti in autonomia dagli allievi della Scuola.

Si è scelto pertanto di definire le **linee guida per la gestione di siti** e server tenendo in considerazione la situazione attuale, in modo da rispettare l’architettura esistente e al contempo garantire stabilità e sicurezza all’intero sistema.

Nel seguito sono descritte nello specifico le proposte di azione della Scuola in base alle varie fasi del ciclo di vita dei siti web.

IL CICLO DI VITA DI UNA WEB APPLICATION

Si definisce **ciclo di vita di un'applicazione web** l'intera esistenza di un sito/applicazione web: dalla sua pubblicazione all'eventuale spegnimento.

In questo ciclo di vita si identificano dei momenti particolari per i quali sono state descritte le proposte di azione:

- la nascita: attivazione
- la maturità: il sito è pubblico, funzionante e mantenuto
- l'obsolescenza: chiusura, reingegnerizzazione, staticizzazione o limitazione alle funzionalità.

LA NASCITA: ATTIVAZIONE

Possono richiedere l'attivazione di un sito i responsabili di progetto, i responsabili dei laboratori, centri, gruppi di ricerca, i professori e i ricercatori.

Al momento della attivazione di un nuovo sito o applicazione web la procedura indicata prevede una **richiesta di attivazione** rivolta all'Area Strategie Digitali (Servizio Infrastrutture Informatiche e Servizio Sistemi Informativi) in cui specificare le seguenti informazioni:

- dati del responsabile scientifico
 - dati delle figure di amministratore del sistema e dei dati
 - progetto istituzionale a cui è collegato il nuovo sito e sua durata
 - nome del sito
 - elenco tecnologie utilizzate
 - sistema operativo su cui installare il sito
 - eventuali librerie o estensioni da attivare/installare
 - indicazione di massima dell'occupazione disco prevista per il sito
 - indicazione del volume di traffico previsto
- richieste ulteriori: attivazione di un sistema di statistiche, ecc.

L'ASD renderà disponibile la documentazione e i materiali di supporto alla procedura di attivazione del sito come ad esempio documenti relativi alle *best practices* per sviluppare in sicurezza e per permettere la coerenza delle nuove installazioni nonché il presente documento per descrivere le procedure da seguire.

Nel caso di richieste di utilizzo di tecnologie o librerie/estensioni non standard o proprietarie, è necessario analizzare eventuali problemi legati alla sicurezza, alla compatibilità e manutenibilità del nuovo oggetto da installare rispetto alle caratteristiche dell'infrastruttura SNS.

Nel caso di richiesta di utilizzo di tecnologie non compatibili, saranno proposte alternative supportate in quanto standard o già esistenti nell'infrastruttura SNS. A questo fine sarà definito un apposito elenco delle tecnologie considerate *safe*.

Nel caso di richiesta di soluzioni sviluppate ad-hoc o per cui non sia possibile proporre l'utilizzo di soluzioni standardizzate (es. implementazione di un componente o software sperimentale) si richiede, almeno per la

durata del progetto/laboratorio, **l'aggiornamento della documentazione e del codice sorgente che dovranno essere sempre reperibili** attraverso appositi sistemi di versioning.

La fase di attivazione è considerata terminata positivamente con l'accesso all'ambiente di sviluppo del sito e rilascio di nome, IP, e account.

Il richiedente a questo punto è realizza in autonomia il proprio sito, garantendone la sopravvivenza attraverso la procedura di rinnovo, come descritto nel seguito.

LA MATURITÀ: SITO PUBBLICO, FUNZIONANTE E MANUTENUTO

In questa fase le applicazioni web possiedono i requisiti di sicurezza, affidabilità e manutenibilità richiesti dalla Scuola: sono quindi in uno stato di funzionamento ordinario.

Per garantire che questi siti restino nello stato di funzionamento ordinario, si prevede che ogni due anni sia eseguita **la procedura di rinnovo** dell'hosting.

La procedura di rinnovo prevede una **verifica** delle seguenti condizioni:

- permanenza di un responsabile scientifico;
- permanenza di un responsabile tecnico;
- corretto e costante aggiornamento;
- sicurezza delle tecnologie utilizzate.

Un mese prima dello scadere del biennio responsabile scientifico sarà chiamato a confermare o meno la propria intenzione a mantenere in vita il sito, e in caso positivo, indicare il responsabile tecnico. In assenza di responsabile scientifico, il sito non potrà essere mantenuto attivo, a meno che la Scuola non proceda ad una nuova nomina.

Il responsabile scientifico dovrà individuare un responsabile tecnico, confermando quello precedentemente nominato o indicandone uno diverso.

Successivamente alla verifica della presenza dei responsabili scientifico e tecnico, sarà svolta la verifica tecnica ai fini della sicurezza del sistema. Se il sito **non** dovesse superare la verifica tecnica, sarà necessaria una reingegnerizzazione del sito o un aggiornamento delle tecnologie utilizzate per riportare il sito in una condizione di sicurezza. Se il responsabile tecnico non dovesse provvedere tempestivamente alla messa in sicurezza del sito, saranno attuate le indicazioni del paragrafo *“Chiusura, reingegnerizzazione, staticizzazione o limitazione alle funzionalità”*.

Nel caso in cui, alla scadenza dei due anni, non fosse possibile ottenere risposta da parte dei responsabili designati, si procederà a rendere il sito non più accessibile dall'esterno, per un periodo indicativo di sei mesi, rinnovabili in caso di valutazione della possibilità di sostituzione del responsabile scientifico. Nel caso in cui la situazione non trovi una risoluzione, dovranno essere attuate le indicazioni del paragrafo *“Chiusura, reingegnerizzazione, staticizzazione o limitazione alle funzionalità”*.

Tutte le applicazioni web, come qualsiasi altro prodotto software, sono destinate a divenire obsolete e insicure dopo pochi anni. L'obiettivo di questo documento è garantire un elevato livello di sicurezza dell'intera rete della Scuola offrendo indicazioni per mantenere in vita siti web anche di realizzazione non recente.

Per **chiusura** si intende l'interruzione del servizio, l'eventuale ri-assegnazione dell'indirizzo IP e il mantenimento su server del codice e delle eventuali altre risorse utilizzate (db, immagini, ecc.) per il periodo di un anno, eventualmente rinnovabile, al termine del quale non sono più garantiti gli standard di conservazione e backup previsti.

Analizzata la situazione attuale della Scuola, si possono presentare i seguenti casi:

1. il sito web non ha più responsabili scientifico e tecnico
2. la funzione del sito web è terminata (es. conclusione del progetto)
3. il sito web ha almeno un responsabile e:
 - a. è obsoleto e non più aggiornabile
 - b. dopo esito negativo della valutazione preventiva, nessuno ha provveduto a eseguire il riallineamento del sito alle nuove tecnologie
4. il sito web è stato violato ed è divenuto pertanto insicuro per l'infrastruttura SNS

Nel **caso 1** la Scuola potrà decidere se nominare un nuovo responsabile scientifico e mantenere attivo il sito o meno (es.: nel caso della chiusura di un laboratorio, la Scuola può decidere di mantenerne comunque visibile la storia).

Nel **caso 2** il sito sarà chiuso, ferma restando la possibilità per il responsabile scientifico, di tenerne memoria in pagine statiche, su server sicuro. .

Nel **caso 3-a** non è più possibile procedere a nuovi aggiornamenti senza compromettere irreversibilmente le funzionalità del sistema. In questa situazione la Scuola si riserva di valutare singolarmente ogni caso (eventualmente nominando una specifica commissione scientifico-tecnica) ed eventualmente provvedere ad una reingegnerizzazione.

La valutazione dovrebbe considerare, tra gli altri, i seguenti parametri:

- numero di accessi al sito
- superamento dei test di sicurezza da eseguirsi tramite software specifici
- livello d'interesse della comunità scientifica al prodotto
- rilevanza all'interno della Scuola

Nel caso di decisione di mantenimento del sito, dovranno essere presi dei provvedimenti tecnici per limitare i rischi derivanti dall'esistenza di un sito obsoleto e garantire i livelli di sicurezza standard della rete della Scuola. Di seguito sono proposte delle possibili soluzioni tecniche per limitare i rischi:

- siti i cui contenuti possono essere resi staticamente: potrebbe essere fatta una copia statica del sito indipendentemente dall'interesse scientifico (**come descritto in "Archiviazione di siti Web: strumenti tecnici, standard internazionali e buone pratiche"**)

- siti non staticizzabili a basso numero di accessi: potrebbero essere resi accessibili solo da rete interna;
- siti non staticizzabili a elevato interesse scientifico: potrebbero essere utilizzate tecnologie per renderli rispondenti il più possibile ai parametri di sicurezza. In base ai risultati del test sulla sicurezza, l'ASD potrebbe ad esempio utilizzare strumenti di analisi del traffico del sito per impedire i più comuni attacchi come:
 - SQL Injection
 - Cross Site Scripting
 - Bad User Agent
 - Sessions Hijacking.

Per questi siti si potrebbe prevedere anche una graduale limitazione delle funzionalità in caso di variazioni significative nel volume di traffico o di sospetti di violazione.

Nel **caso 3-b** in cui il sito sia aggiornabile ma nessuno abbia preso in carico tale compito, si ritiene che una prima possibile azione sia limitare la visibilità del sito alla sola rete interna fino all'avvenuto aggiornamento.

Nel **caso 4** l'unica via possibile è la **chiusura del sito compromesso**, fino ad eventuale risoluzione radicale del problema.

CONSERVAZIONE DI SITI WEB: STRUMENTI TECNICI, STANDARD INTERNAZIONALI E BUONE PRATICHE

CREAZIONE DELLA COPIA STATICA DI UN SITO

L'operazione di copia in pagine statiche HTML di una o più pagine di un sito web è possibile solo in presenza di alcune condizioni. In generale non è possibile effettuare la copia di un servizio web mantenendo le funzionalità di ricerca e le interfacce per la navigazione dei contenuti che sono il frutto dell'inserimento di parametri arbitrari dell'utente e vengono processate da un algoritmo sul server (CGI) interrogando un database, ma è utile per tutte o quasi le altre situazioni e consente di effettuare copie con pagine statiche non modificabili, se non editando direttamente il codice HTML, ma sicure rispetto alle comuni vulnerabilità dei siti web.

IL FORMATO WARC

WARC (Web ARChive file format) è il formato per l'archiviazione dei contenuti del World Wide Web definito dallo standard internazionale ISO 28500. Il formato è stato sviluppato grazie ai contributi delle principali istituzioni che hanno la missione di preservare nel tempo i siti web. Fra tutte è il caso di ricordare in particolare l'Internet Archive¹ per l'importanza della sua iniziativa e per il fatto che il formato WARC è derivato dal formato ARC_IA² realizzato proprio da questa istituzione.

¹ <https://archive.org>

² <http://www.digitalpreservation.gov/formats/fdd/fdd000235.shtml>

Con WARC è possibile raggruppare molteplici risorse digitali con metadati ed informazioni relative a queste risorse in un unico file di archiviazione. La standardizzazione del formato garantisce la durata nel tempo dell'archiviazione Web e l'interoperabilità fra archivi e collezioni di diverse istituzioni.

Nel definire il formato del file WARC lo standard ISO 28500:2009 individua metodi e convenzioni per

- conservare le richieste (per esempio i request header) oltre che i risultati per i principali protocolli sul livello applicativo di Internet: Hypertext Transfer Protocol (HTTP), Domain Name System (DNS) e File Transfer Protocol (FTP);
- conservare arbitrari metadati da associare ai dati archiviati (classificazioni per soggetto, lingue utilizzate, encoding, etc.);
- supportare la compressione dei dati mantenendo l'integrità;
- conservare tutte le informazioni di controllo dal protocollo;
- conservare gli eventi duplicati come link a dati già presenti nell'archivio (per ridurre le necessità di storage in presenza di risorse identiche o assimilabili);
- garantire future estensioni delle funzionalità del formato standard senza perdere le funzionalità esistenti.

Gli standard ISO non sono consultabili gratuitamente, le sezioni informative sono pubbliche ma il full text è disponibile solo a pagamento³, è però possibile consultare liberamente la bozza finale dello standard ISO⁴.

SISTEMA DI GESTIONE DEI SITI ISTITUZIONALI DI CENTRI/LABORATORI E GRUPPI DI RICERCA

Per quanto riguarda i nuovi siti, l'Area Strategie Digitali propone un protocollo specifico dedicato ai siti istituzionali di centri, laboratori e gruppi di ricerca.

Si propone una soluzione standardizzata che prevede l'utilizzo di un sistema appositamente progettato, sviluppato e gestito nell'ambito della progettazione del nuovo portale SNS, basato su Drupal, per la realizzazione e gestione in autonomia del proprio sito web con una grafica in linea con quella della Scuola.

L'ASD propone una soluzione che consenta la realizzazione e la gestione in autonomia del proprio sito web mantenendo/preservando l'allineamento con gli stili grafici del sito principale della Scuola, secondo le indicazioni fornite dagli uffici preposti.

Il sistema offre una soluzione che consente di integrare in modo efficiente e efficace i contenuti inerenti l'ambito della ricerca che sono presenti in altri sistemi informativi della Scuola, come i moduli U-GOV.

Vengono offerti di default tre template diversi, ma sempre in linea con lo stile del sito istituzionale, completamente responsive e quindi compatibili con tutti i dispositivi esistenti (pc, tablet, mobile...). Le personalizzazioni previste riguardano:

³ <https://www.iso.org/obp/ui/#iso:std:iso:28500:ed-1:v1:en>

⁴ http://bibnum.bnf.fr/warc/WARC_ISO_28500_version1_latestdraft.pdf

- i colori: scelta tra una vasta gamma di palette di colori e font SNS per la progettazione di un look unico per il sito
- la testata: inserimento del logo
- le voci di menu.
- Perché viene proposto questo sistema:
 - facilità di utilizzo: permette di concentrarsi su ciò che conta di più: sul sito web. Attraverso un editing semplificato permette di: aggiungere nuovi contenuti, aggiornare il contenuto, gestire gli utenti e gli editor di contenuti e menù;
 - mobile-friendly e accessibile
 - integrazione con i sistemi della Scuola: perché aggiornare lo stesso contenuto in più di un luogo? Questo sistema si connette ai dati della Scuola, semplificando la manutenzione delle informazioni
 - non servono competenze tecniche elevate
 - gestione centralizzata ma diversificata per livelli di utenza
 - supporto per la configurazione a carico dell'ASD.

I contenuti previsti possono essere personalizzati sulla base di una proposta generale che prevede:

- indirizzi di ricerca
- eventi
- pubblicazioni
- persone
- progetti

INDIRIZZI DI RICERCA

Si intendono le aree/ambiti di ricerca dello specifico laboratorio, come

- nome
- immagine/logo
- abstract/testo descrittivo
- contatti
- responsabile dell'attività di ricerca (potrebbe essere una relazione verso la persona, dando per scontato che siano persone interne alla Scuola, o comunque presenti in U-GOV)
- informazioni redazionali relative al responsabile

Oltre ai campi specifici, nella pagina dell'indirizzo sono disponibili, dinamicamente:

- l'elenco dei progetti classificati per tale indirizzo
- l'elenco delle persone afferenti
- l'elenco degli eventi classificati per tale indirizzo

EVENTI

Si intendono gli eventi dello specifico laboratorio nel relativo sito. In questa fase questi contenuti sono gestiti redazionalmente. Oltre ai campi che assolvono le esigenze di comunicazione della Scuola, è presente

un campo che referencia uno o più indirizzi di ricerca (per consentire l'esposizione degli eventi nelle pagine degli specifici indirizzi) ed un campo redazionale che referencia uno o più progetti (per consentire l'esposizione degli eventi nelle pagine degli specifici progetti).

PUBBLICAZIONI

Le pubblicazioni sono acquisite da IRIS e gestite con un entità specifica. Per tale entità non saranno previsti campi redazionali.

I campi minimi per garantire la ricerca sono:

- titolo
- autori
- anno
- progetto correlato
- indirizzo di ricerca correlato.

Il risultato della ricerca espone tutte le pubblicazioni associate ad almeno un progetto pubblico presente sul sito del laboratorio.

Tra i campi è presente anche il link alla specifica pubblicazione in IRIS (pagina pubblica della pubblicazione) e l'elenco delle pubblicazioni rimanderà alla pagina di dettaglio in IRIS.

Questi campi minimi consentono di poter usare l'entità delle pubblicazioni come oggetto referenziabile dentro agli altri content type o altre entità.

PERSONE

Le persone sono estratte direttamente da U-GOV e gestite con un entità specifica. Per tale entità non sono previsti campi redazionali.

I campi minimi per garantire la corretta gestione sono:

- cognome,
- nome,
- e-mail,
- identificativo,
- url alla scheda delle rubrica/esposizione dei dati della rubrica.

Oltre ai campi specifici, nella pagina della persona sono disponibili, dinamicamente:

- l'elenco dei progetti di cui è responsabile o a cui partecipa
- l'elenco degli indirizzi a cui afferisce
- l'elenco delle pubblicazioni di cui è autore/autrice

Da verificare con le singole strutture richiedenti se inglobare la scheda personale presente nel portale istituzionale o inserirne una diversa, specifica per la persona nel contesto del laboratorio.

PROGETTI

Per la gestione dei progetti si propone una soluzione IBRIDA:

i progetti derivano parte dei campi da U-GOV PJ e parte da un'entità gestibile redazionalmente. Sono quindi previsti i campi semplici e più strutturati presenti dentro U-GOV PJ e i seguenti campi redazionali:

- immagine,
- url del progetto,
- contatti,
- relazione verso uno o più indirizzi di ricerca (entità),
- relazione verso uno o più pubblicazioni (entità),
- relazione verso una o più persone (entità): responsabili, come già censiti in U-GOV PJ, e persone attualmente referenziate in SerSe (da passare automaticamente in PJ)
- elenco di persone non mappate dentro all'entità con un elenco di campi minimi (es. cognome, nome, email, url sito web)

Questo strumento è a disposizione del personale docente e utilizzabile previa richiesta al Servizio Sistemi Informativi.

ATTUAZIONE DELLE LINEE GUIDA A UN CASO DI STUDIO: CHIUSURA DEL CENTRO DI FILOSOFIA

Il caso di studio ha riguardato il Centro di Filosofia che ha terminato la sua attività di ricerca e ha lasciato in eredità alla Scuola diversi siti di progetto interni e in conto terzi, alcune sperimentazioni web e il sito del centro stesso; il responsabile dei progetti esiste e ha confermato la sua volontà di mantenere vivi questi siti che si sono rivelati in parte non più aggiornabili.

In particolare i siti sono:

1. bivio.filosofia.sns.it
2. bibliotecaideale.filosofia.sns.it
3. bibbia.filosofia.sns.it
4. rarebooks.filosofia.sns.it
5. imagohistoriae.filosofia.sns.it
6. vitadibruno.filosofia.sns.it
7. textandsemantic.filosofia.sns.it

Le azioni svolte sono state di due tipi: per i siti aggiornabili (da 4 a 7) si è provveduto al trasporto sui server ASD, reinstallando i back-up dei database e copiando i siti con le nuove configurazioni.

Per i siti da 1 a 3, che utilizzano un software realizzato ad hoc, non più sviluppato e legato al server su cui gira si è deciso di prendere l'intera macchina virtuale KVM e copiarla sui server fisici gestiti da ASD.