

Indicazioni sulla politica di gestione dei dati prodotti dal PNRA

A cura del GdL *‘Raccolta diffusione dati del PNRA’*

Stefano Aliani, Lili Cafarella, Guido Di Donfrancesco, Antonio Meloni, Giuseppe Orombelli, Nigel Wardell

Presentazione

Questo documento è il risultato di incontri e scambi di email da parte del GdL *‘Raccolta e diffusione dati del PNRA’* in risposta all’obiettivo stabilito dal MIUR all’art 1 del decreto istituzionale del gruppo:

“E’ strategico che il PNRA costituisca un sistema di raccolta e diffusione dei dati prodotti sia nell’ambito dei progetti di ricerca, sia da parte degli osservatori permanenti.

In particolare dovrà essere realizzato un portale dei dati PNRA e dovranno essere costituiti i competenti data centers.

Il compito del gruppo di lavoro è quello di predisporre, entro 6 mesi, una proposta di regolamento che riguardi la policy relativa sia al rilascio, la raccolta e la conservazione dei dati sia all’accesso agli stessi da parte dei gruppi di ricerca nazionali ed internazionali.”

Alle riunioni svolte al MIUR e al CNR hanno partecipato in qualità di consulenti anche esperti coinvolti in attività di interesse del GdL che hanno contribuito ad approfondire aspetti tecnici relativamente al problema dati del PNRA.

L’importanza del libero e pieno accesso ai dati originali, di cui i Metadati sono la chiave di ingresso, è fondamentale nella moderna ricerca perché il permettere alla comunità scientifica di accedere alla enorme mole di dati raccolti dai predecessori, è alla base del progresso scientifico; consente lo sviluppo della buona scienza e consente di prendere decisioni per le attività future. Inoltre la condivisione dei dati aiuta i ricercatori a ottenere credito per il loro lavoro ed aumenta la propria visibilità. Quando i dati sono ben gestiti, sono più facili da trovare, da capire e da utilizzare. L’accessibilità aiuta a spostare in avanti la frontiera della scienza, così che tutti possono sfruttare il lavoro fatto negli anni passati e farlo diventare parte del loro lavoro.

Il presente documento riporta le azioni intraprese e quelle da effettuare per la realizzazione di una politica italiana relativamente ai dati prodotti dal PNRA, considerando la attuale strutturazione del Programma, i recenti sviluppi tecnologici nel campo della raccolta, conservazione e condivisione dei dati.

Le indicazioni che il GdL ha pensato di aggiungere sono finalizzate a:

- Garantire la **stabile disponibilità** dei prodotti dal PNRA, rendendoli sempre fruibili per la ricerca e la divulgazione scientifica;
- Agevolare la **diffusione dei dati** prodotti dal PNRA verso la comunità scientifica nazionale e internazionale;
- Garantire la **qualità e la tracciabilità dei dati** raccolti nell'ambito di progetti finanziati dal PNRA.

Le informazioni riportate in questo documento, dopo una breve introduzione sullo stato dell'arte, descrivono:

1. Breve definizione di dato e metadato
2. Descrizione del ruolo del National Antarctic Data Center (NADC)
3. Vincoli e accordi tra i proponenti di ricerca e PNRA
4. Conferimento dei dati al NADC
5. Gestione condivisione e pubblicazione dei dati
6. Il diritto d'autore
7. La politica internazionale dei dati e lo SCAR.

Lo Stato dell'Arte

La direttiva europea sul riutilizzo dell'informazione del settore pubblico (direttiva 2003/98 / CE, conosciuta come la 'direttiva PSI') è entrata in vigore il 31 dicembre 2003. Essa incoraggia gli Stati membri a rendere quante più informazioni disponibili in maniera aperta e regolamentata. Si rivolge a tutto il materiale informativo detenuto da enti del settore pubblico negli Stati membri, a livello nazionale, regionale e locale (quali ministeri, enti pubblici, comuni), così come le organizzazioni finanziate per la maggior parte da o sotto il controllo delle autorità pubbliche (ad es. istituti meteorologici). Il regolamento successivo, 1367/2006/EC13, è molto simile alla direttiva 2003/4/CE, ma riguarda l'accesso alle informazioni ambientali detenute dalle "istituzioni e organismi comunitari". Questo regolamento completa la direttiva 2003/98/CE "sul riutilizzo dell'informazione del settore pubblico" che tratta l'accesso e il riutilizzo dell'informazione in questo settore.

La direttiva è stata recepita in Italia con il [Decreto legislativo del 24 gennaio 2006, n. 36](#), entrato in vigore il 1 marzo 2006.

Secondo le direttive dello SCAR la data policy generale dei progetti antartici si deve basare sul presupposto che tutti i dati raccolti dai vari Programmi internazionali e nazionali, e quindi anche dal PNRA, debbano essere liberamente disponibili e

vengano distribuiti apertamente per scopi didattici, scientifici e non commerciali in modo e tempi opportuni a salvaguardare eventuali diritti d'autore.

E' pertanto indispensabile che il PNRA segua queste indicazioni e renda obbligatorio applicare policy dei dati conformi agli accordi internazionali a tutti i dati prodotti dai progetti finanziati dal PNRA a partire dalla campagna 2014-15 e, per quanto possibile, attivare azioni per cui anche i dati raccolti in passato, seguano lo stesso trattamento.

Dopo quasi trenta anni di attività, Il PNRA ha un enorme mole di dati acquisiti e la loro gestione è sempre stata una parte rilevante del Programma.

Già nel 1992 sotto gli auspici dello SCAR e dei mandati dell'Antarctic Treaty Consultative Meeting Recommendation XVI-12, è stato attivato un sistema di condivisione di dati geofisici curato all'inizio dall'USGS, e confluito poi nell'OGS nel 1997, con il sostegno del PNRA. L'SDLS (Antarctic Seismic Data Library System for Cooperative Research) è una banca dati progettata da membri della comunità geoscientifica Antartica (programma ANTOSTRAT) per facilitare l'accesso aperto ai dati sismici multicanali (MCS) di riflessione, e di promuovere la ricerca collaborativa sulla base di tali dati. Attualmente sono disponibili dati e informazioni relative a 250000 km di linee sismiche a mare provenienti dai progetti italiani ed internazionali in Antartide. Nel tempo diverse istituzioni scientifiche e gruppi di lavoro operanti nel PNRA hanno realizzato banche dati tematiche, ma senza un'adeguata azione di coordinamento a livello nazionale.

Nel 2003 è stato avviato, come primo passo verso la gestione dei dati, il Progetto Sistema Informativo Ricerca Italiana in Antartide (SIRIA), coordinato dal CNR che mirava alla gestione dei Metadati delle attività di campagna. Il programma SIRIA ha raccolto nel periodo 2004-2009, 137 metadati provenienti dai progetti italiani, organizzati secondo format specifici per ciascuna area disciplinare (vedi appendice A).

La recente ristrutturazione del PNRA del 2010, ha individuato nel CNR l'attuatore dei Piani Esecutivi Annuali e il Responsabile della diffusione e gestione dei dati scientifici.

1) Definizione del dato e del metadato

Un **dato** comprende ogni tipo di osservazione scientifica di dettaglio, sia essa originale che processata, che consista in una informazione scientifica auto-consistente archiviabile, dalla forma cartacea a quella digitale. Il dato è, nel nostro ambito, il fine primo di una osservazione scientifica elementare. I dati possono presentarsi sotto diverse forme: numeri (ad esempio: temperatura, umidità, spessore del ghiaccio, coordinate geografiche, etc...) e [testo](#), come nel caso di una descrizione di un animale, un paesaggio, un fenomeno. Ma ovviamente sono dati anche [immagini](#) statiche, immagini in movimento ([video](#)), [suono](#) ([audio](#)) ed altro ancora. I dati

possono essere conservati su diversi supporti fisici e/o veicolati attraverso una [rete di telecomunicazione](#) tra più utenti.

Un **metadato** è invece un tipo di informazione che descrive il dato e consente il raggiungimento e la disponibilità del dato di interesse. In particolare il metadato contiene ad esempio le seguenti informazioni:

- Individuazione e localizzazione di un documento contenente i dati;
- [Interoperabilità](#) semantica, consistente nel permettere la ricerca in ambiti disciplinari diversi che può essere ottenuta grazie a una serie di equivalenze fra [descrittori](#), per condurre all'ottenimento del dato di interesse.
- La gestione risorse, ossia gestire le raccolte di documenti grazie all'intermediazione di banche dati e cataloghi;
- La disponibilità, ovvero ottenere informazioni sull'effettiva disponibilità del documento.

Quindi i metadati rappresentano lo strumento per facilitare l'individuazione, l'accesso e l'uso pratico dei dati stessi. E' 'obbligo' delle nazioni aderenti allo SCAR di rendere disponibili i metadati delle proprie ricerche in Antartide presso l'AMD (Antarctic Master Directory) appositamente realizzata.

Tipologie di dati:

I dati prodotti dal PNRA possono essere a grandi linee separati in due classi: dati di Osservatori (permanenti o temporanei) e dati raccolti nell'ambito di progetti di ricerca

a) Dati di osservatorio:

I **dati degli osservatori** sono in generale facilmente individuabili perché nella tipologia della ricerca associata all'Osservatorio, la produzione del dato ne costituisce l'essenza. Ne sono esempi gli osservatori meteo-climatologici, quelli sismologici, oceanografici, etc... Questi dati si archiviano digitalmente. I metadati di questo tipo di dati sono piuttosto semplici e schematici: devono comprendere l'identificazione della località dove l'osservazione avviene, le caratteristiche della grandezza misurata, il tempo di campionamento, la precisione delle misure effettuate.

E' auspicabile che i dati di osservatorio siano messi a disposizione in tempo reale o quasi, compatibilmente con i mezzi telematici messi a disposizione dalla logistica che opera nelle stazioni antartiche. Gli stessi dati dovrebbero poi essere trasferiti a database internazionali per consentirne anche una visibilità diversa (anche su questo si veda in seguito).

E' necessario inoltre distinguere tra osservatorio e **sistema osservativo**. Quando la stazione osservativa fornisce solo dati su certi periodi temporali limitati non può

essere considerata osservatorio ma bensì sistema osservativo. E' questo il caso ad esempio di molta della strumentazione operante in astrofisica che, pur se stanziale non ha caratteristiche di osservazione continuativa e quindi conseguente produzione di dati con continuità.

b) Dati da progetto

Le attività di progetto sono di tipo non ripetitivo ed esplorativo. In questo caso il metadato, che dovrebbe contenere informazioni sulla localizzazione geografica dei siti raggiunti, riporta informazioni sul tipo di attività svolta, in modo da contribuire anche a segnalare aree eventualmente alterate o contaminate. Pur essendo stata indicata anche per i progetti 'ordinari' una 'policy' nell'ambito della fase di proposte di ricerca, non si può dire che il PNRA abbia stabilito, al momento, al riguardo un 'vero' protocollo o un accordo in proposito. In ogni caso il PNRA non è stato in grado di metterlo in atto 'facendolo rispettare'. Appare quindi necessario stabilire con i proponenti, un accordo al riguardo che, per le differenze intrinseche di tipologia del dato, è diverso da quello richiesto per gli osservatori.

Trattando la tipologia di dati di progetto, inevitabilmente si finisce con il coinvolgere anche i prodotti di ricerche che non rientrano 'ordinariamente' nella tipologia del dato. E' il caso generico dei 'reperti', oggetto di discussione di un altro GdL. La responsabilità dei reperti è del Museo dell'Antartide. Da qualche anno i reperti raccolti passano dal luogo di prelievo (ad esempio la stazione antartica) direttamente al ricercatore, creando una situazione di possibile confusione rispetto a parametri di campagna (luogo e tempo di prelievo, etc...) vitali per la ricerca. E' necessario coordinare sia la programmazione prima della campagna che la distribuzione successiva, in modo da poter tracciare in ogni momento la posizione dei campioni raccolti. Nell'attuale organizzazione questo lavoro va fatto in concerto tra Museo, CNR ed ENEA.

2) Il Data Center Antartico Nazionale

Per garantire la **stabile disponibilità**, in conformità con gli indirizzi forniti dal Decreto di Riordino del PNRA, è costituito presso CNR DTA un sistema di gestione dei dati e metadati antartici denominato attualmente [*Italian Polar Metadata System*](#), che costituisce la base del Data Center Antartico Nazionale (National Antarctic Data Center Italiano, NADC) come previsto da report della commissione SCADM dello SCAR.

http://scadm.scar.org/0files/SCAR_DIMS_34.pdf

Il sistema si basa sul moderno concetto di interoperabilità delle risorse, principio attraverso il quale banche dati, software, servizi (anche tecnologicamente diversi) possono scambiarsi contenuti e dati attraverso standard condivisi.

Il sistema ha già importato al suo interno i metadati contenuti in SIRIA, sotto forma di un catalogo “storico” di metadati conforme allo standard ISO 19115.

Lo schema più completo dell’infrastruttura proposta con il piano di sviluppo e gestione sono descritti nell’**appendice B**. Le specifiche di dettaglio e tutti i documenti tecnici necessari saranno messi a punto in tempo utile per le prossime chiamate per proposte di ricerca.

Alcune istituzioni (per es. OGS, CNR, INGV, MNA) i cui ricercatori lavorano in Antartide da molto tempo hanno già costituito delle banche dati tematiche disciplinari e potrebbero in tempi brevi diventare parte integrante di uno o più nodi federati.

Governance del Sistema

Il CNR è responsabile dello sviluppo della manutenzione e della operatività del sistema. Un nodo di I livello è responsabile di stabilire gli interoperability agreement con il Common Node e di interfacciarsi con i nodi di livello II, III, ecc.

Comitato di controllo

Ha il compito di garantire la sostenibilità economica della infrastruttura e quindi ripartire le risorse finanziarie. Inoltre il comitato raccoglie i bisogni della comunità e verifica l’efficacia con la quale l’infrastruttura risponde agli stessi.

Il comitato ha il compito di fornire annualmente agli organi di Governance del PNRA (CNR, CSNA) un report sullo stato del sistema.

Il GdL propone che il comitato di controllo sia costituito da un responsabile dell’infrastruttura, un responsabile per il nodo primario (Common Node), i responsabili dei nodi di primo livello (INGV, CNR, OGS, ENEA, MNA), un rappresentante CRP e un rappresentante CSNA.

Sostenibilità

Il PNRA attraverso il PEA deve destinare annualmente risorse al sostentamento dell’infrastruttura del common node e dei nodi di primo livello.

3) Vincoli e accordi tra i proponenti dei progetti di ricerca e PNRA

Alcune indicazioni generali di policy a cui il Data Center Antartico Nazionale e i ricercatori afferenti e non-afferenti al PNRA devono attenersi, sono riportate di seguito.

Affinché il sistema di NADC possa essere pienamente operativo, è necessario che i ricercatori rendano i loro dati accessibili nell’ambito di un chiaro regolamento di uso dei dati del PNRA.

Di seguito sono riportate alcune regole generali da adottare in ambito PNRA che mirano a seguire gli indirizzi generali previsti dall’Unione Europea e dal MIUR per

una logica di *'open access'*.

Gli **accordi** attuativi di dettaglio che saranno stipulati per specifiche tipologie di dati tra Ricercatore e PNRA attraverso il NADC, devono rispettare queste regole generali.

Il GdL ha identificato le seguenti **indicazioni generali** da ritenersi fondamentali per la data policy generale del PNRA e come tali devono essere accettate dai proponenti e dal sistema PNRA, prima dell'approvazione di un Progetto di Ricerca;

1. Il GdL ritiene che sia necessario realizzare, ove possibile, la **raccolta dei metadati** durante le Spedizioni. Nel sistema distribuito riportato nell'appendice B, questo implica di realizzare specifici strumenti informatici sui quali, attraverso apposite procedure, compilare il metadato.
2. Il GdL ritiene che debba essere diffusa ed accettata da tutti **la politica generale che i dati devono essere resi disponibili** a tutta la Comunità scientifica, in prima battuta in ambito nazionale, e quindi internazionale (SCAR). Il ricercatore ha diritto a un periodo di **uso esclusivo** durante il quale deve essere garantita la riservatezza e la disponibilità ristretta dei dati per il ricercatore che li ha prodotti.
3. Il GdL ritiene che debba essere incoraggiata anche **l'archiviazione dei dati** sia nella **forma elaborata** sia in quella più **grezza** e meno elaborata possibile. In questo modo si rende disponibile l'informazione originale, per possibili rianalisi future. Per questo tipo di dati sarà prevista una specifica licenza di uso e condivisione dei dati (si veda l'appendice C - Licenza Creative Commons)
4. Prevedere un **sistema automatico di controllo** per verificare se sono stati soddisfatti gli obblighi del referente del progetto in materia di produzione dei metadati e messa a disposizione dei dati secondo i principi che rispondano agli standard concordati in fase di proposta. Al Responsabile del progetto sarà chiesta spiegazione del mancato conferimento dei dati e gli esiti potranno essere vincolanti nella fase di presentazione di nuove proposte.
5. Programmare il recupero e la catalogazione dei **dati storici** raccolti nell'ambito di passate spedizioni. Per ragioni di praticità, questo punto deve essere affrontato dopo che i nodi del Data Center Antartico Nazionale saranno operativi, ma non deve essere trascurato.

4) Conferimento dei dati al NADC e vincoli relativi

Il GdL ha identificato le seguenti **indicazioni generali** perché diventino regolamento comune del PNRA e siano incluse nei bandi dei progetto di ricerca e

accettate dai proponenti prima della sottomissione delle proposte di ricerca. L'accettazione è vincolante all'approvazione del Progetto e la stesura del Contratto di Ricerca.

1. Il proponente riconosce e accetta che i dati prodotti nella ricerca di cui è responsabile sono di **proprietà del PNRA**, fermi restando i diritti del ricercatore sulla base delle Leggi e dei Regolamenti sulla proprietà intellettuale e gli eventuali Accordi tra le Parti.
2. Il proponente deve sottoscrivere un **accordo di Licenza** d'uso dei dati con il PNRA.
3. Ogni proposta di ricerca deve obbligatoriamente descrivere un **piano di gestione** dei dati (che potrebbe contribuire al punteggio di valutazione complessiva del progetto). Il PNRA valuterà l'opportunità di inserire all'interno del budget del progetto proposto una eventuale **voce di spesa** specifica.
4. In fase di presentazione del progetto, si può prevedere la possibilità di mettere a disposizione i dati, entro 2 anni o alla fine del progetto, in tre modi differenti:
 - a. **Incaricare il NADC** di gestire i dati del proprio progetto
 - b. **Inviare i dati ad un data center certificato** che normalmente si occupano della gestione di quel tipo di dati, garantendo comunque l'accessibilità e l'interoperabilità con il NADC.
 - c. **Gestire direttamente i dati** su server dell'Istituzione di appartenenza conseguentemente ad un impegno formale dell'ente di appartenenza ed osservando le regole di **interoperabilità** stabilite con il NADC.

I dati devono comunque essere conferiti entro un **periodo di tempo adeguato** (che suggeriamo possa essere 5 anni dalla fine del progetto) al NADC o ai suoi nodi (vedi in seguito). I metadati invece devono essere conferiti al NADC al momento della loro creazione, nell'ambito della campagna di misura.

5) Gestione, condivisione qualità e pubblicazione dei dati

Il NADC, insieme ai suoi nodi, ha anche lo scopo di **assistere il singolo progetto o gruppo di ricerca** nelle fasi successive alla raccolta dei dati, inclusa la gestione e condivisione dei dati.

Questo fornisce notevoli vantaggi:

- a. Permette di lavorare con standard e formati condivisi;
- b. Favorisce l'adesione dei ricercatori alla politica open source garantendone i diritti attraverso una logica di licenze di distribuzione agevolando la collaborazione scientifica.
- c. Permette al singolo gruppo ricercatore di ottimizzare i protocolli di

- d. Riduce rischi etici insiti nel portare i dati lontano da chi ne conosce il valore e le problematiche;
- e. Potenzia le strutture di data center già esistenti federandole tra di loro.
- f. **Il Comitato di Controllo del NADC cura una collana di Data Report** fornita di DOI o ISBN (o altro sistema di identificazione bibliometrica) che raccoglie i dati PNRA con la collaborazione dei ricercatori.

Al ricercatore spetta il controllo della conformità e della qualità del dato (QA e QC, Quality Assesment e Quality Control). I data center federati e specializzati potranno anche effettuare dei controlli specifici per il tipo di dati.

6) Il diritto d'autore e i diritti del programma finanziatore

In questo paragrafo sono fornite indicazioni generali sulla condivisione e l'uso dei dati.

- 1) La **proprietà dei dati** rimane del PNRA, che ha l'obbligo di conservarli in modo permanente. Il ricercatore deve conservare i diritti di proprietà intellettuale previsti dalla Legge.
- 2) **L'autorizzazione all'uso** dei dati avviene sempre attraverso l'assegnazione di Licenze, usando il più possibile Licenze del Tipo Creative Commons (CC - <http://www.creativecommons.it/>), ampiamente usate per il mondo Open Source. Accordi dedicati stipulati tra PNRA e ricercatore sono comunque accettabili per situazioni speciali (es. dati sensibili o riservati).
- 3) A ogni data set vengono associate una o più **tipologie di licenza** in base al livello di protezione richiesto dal tipo di dato e dal caso specifico.
- 4) Il responsabile del Progetto di Ricerca sottoscrive l'**Accordo di Licenza** durante la preparazione delle Proposte di Ricerca.
- 5) Chi **deposita** i dati deve aver sottoscritto la Licenza che ne controlla l'uso e la distribuzione secondo gli accordi di data policy stabiliti tra Ricercatore e PNRA in fase di stesura della Proposta di Ricerca.
- 6) Chi **richiede** i dati è sottoposto ai vincoli d'uso e distribuzione stabiliti dalla Licenza che sottoscrive al momento della richiesta di accesso ai dati. Le licenze sono da intendersi valide sia per richiedenti nazionali sia internazionali.
- 7) **I dettagli dei termini** di ciascuna tipologia di Licenza e le modalità con le conseguenti regole per l'autorizzazione all'uso dei dati sono definite dal Comitato di Controllo del Data Center Antartico Nazionale, sentiti i ricercatori competenti. I data center esterni al PNRA devono sottoscrivere opportuna licenza i cui termini saranno da definire di caso in caso.
- 8) In ogni circostanza in cui sono utilizzati dati PNRA si richiede il **riconoscimento formale**, qualora non siano tra gli autori, di chi ha contribuito alla raccolta e gestione dei dati, oltre a riferimenti precisi alle fonti di finanziamento che ne hanno permesso la raccolta. Dove possibile, questo

riconoscimento dovrebbe assumere la forma di una citazione, similmente a ciò che si fa ad esempio citando un articolo di libro o una rivista.

- 9) Gli utilizzatori dei dati devono citare la **collana di Data Report** del PNRA.
- 10) L'uso a **fini commerciali**, o comunque non scientifici o didattici, dei dati è sottoposto a vincoli e limitazioni diverse e sarà trattato caso per caso dagli organismi di gestione del PNRA.
- 11) Molte riviste scientifiche chiedono il **trasferimento del copyright** della pubblicazione e dei dati da cui è derivata. Questo non deve contrastare con le licenze CC e con i regolamenti PNRA. Sia il PNRA sia l'autore dell'articolo devono prestare attenzione a questi aspetti.
- 12) I codici dei modelli utilizzati dai ricercatori non rientrano nella politica di conferimento.

I dettagli degli accordi tra PNRA e Responsabili di progetto sono definiti dal Comitato di Controllo del NADC.

Creative Commons è un'organizzazione non-profit. Le licenze Creative Commons, come tutti i nostri strumenti, sono utilizzabili liberamente e gratuitamente, senza alcuna necessità di contattare CC per permessi o registrazioni.

7) La politica dati internazionale e lo SCAR

Lo SCAR ha al suo interno un comitato responsabile per tutti gli aspetti della gestione dei dati e le relative informazioni: il comitato permanente sulla gestione dati Antartici, lo SCADM (Standing Committee on Antarctic Data Management). CSNA e CRP propongono i nomi dei rappresentanti italiani allo SCADM. Il rappresentante deve regolarmente riferire le decisioni prese dalla comunità internazionale sul data management ma deve anche essere referente della comunità scientifica italiana in ambito SCADM raccogliendo prima dei meeting internazionali informazioni sullo stato dell'arte della situazione italiana.

Inoltre SCAR fornisce (attraverso lo SCADM) un Antarctic Data Management System (ADMS) che sostiene un portale per il riconoscimento delle informazioni sui sistemi dati, l'Antarctic Master Directory (AMD), e conta su un sistema distribuito operante nelle varie nazioni per lo 'storing' e la manutenzione del dato, i National Antarctic Data Centres (NADC).

Nell'Agosto 2014 si è svolta, nell'ambito dello SCAR XXXIII, la riunione dello SCADM dove si è ribadito l'obbligo di mettere on-line i dati originali.

E' necessario che venga resa pubblica la costituzione di un Data Center Antartico Nazionale Italiano strutturato su una logica di data brokering e networking e vengano

organizzati periodicamente meeting ed eventi divulgativi e vengano raccolte le esigenze della comunità scientifica.

RINGRAZIAMENTI

Si ringraziano Simona Longo, Stefano Nativi, Claudio Rafanelli, Vincenzo Romano per la consulenza tecnica gentilmente fornita.

APPENDICE A

Organizzazione di metadato per i dati PNRA nell'ambito del progetto SIRIA

Tematiche di ricerca rappresentati nel PNRA:

1. Biology and Medicine
2. Geodesy and Observatories
3. Geophysics
4. Geology
5. Glaciology
6. Science of the Atmosphere
7. Sun-Earth Relationships and Astrophysics
8. Oceanography and Marine Ecology
9. Chemical Pollution
10. Jurisprudence and Geography
11. Technology

La struttura del progetto SIRIA relativo ai metadati prevedeva il seguente schema di informazioni:

1. Dataset identification
2. Dataset overview
3. Dataset quality elements
4. Spatial reference system
5. Extent
6. Data definition
7. Classification
8. Administrative metadata
9. Metadata reference
10. Metadata language

Standard utilizzato:

STANDARD CEN/ TC 287

I dati del progetto SIRIA sono confluiti nel [*Italian Polar Metadata System*](#).

APPENDICE B

Infrastruttura del NADC e i suoi nodi

Questa appendice ha lo scopo di illustrare brevemente una possibile infrastruttura informatica che garantisca, a livello nazionale, la raccolta e la diffusione dei dati prodotti in ambito polare.

Lo schema che si prevede di utilizzare è finalizzato a raccogliere, conservare, mantenere, pubblicare, rendere accessibile e preservare dalla perdita l'immenso patrimonio di informazioni (dati scientifici) che ogni anno vengono acquisite e prodotte in ambito PNRA.

L'architettura che si propone si basa sul concetto di "sistema dei sistemi" un insieme cioè di sistemi interconnessi tra loro per mezzo di servizi di mediazione, (brokering approach) che consentono di mettere in comune le loro risorse e capacità per dar vita ad un sistema più complesso che offre maggiori funzionalità e prestazioni che non la semplice somma dei sistemi esistenti.

Quest'approccio di tipo "bottom-up" garantisce l'autonomia e le diversità tecnologiche dei singoli sistemi, e al tempo stesso la piena interoperabilità delle risorse.

L'interoperabilità è infatti quell'importante tema e principio che prevede che software, banche dati, servizi, devices e apparati possano scambiare contenuti e dati attraverso standard condivisi, senza che gli utilizzatori debbano affrontare particolari costi/oneri relativi alla conversione dei dati; l'interoperabilità serve a superare quelle situazioni in cui l'utente rimane vincolato all'utilizzo di un certo sistema perché non può o non vuole affrontare una difficile operazione di conversione dei propri dati, alcune volte addirittura impossibile.

Quest'approccio architetturale consente di interconnettere sistemi ed infrastrutture digitali eterogenee in modo flessibile, espandibile e sostenibile. Questi sistemi possono così far parte di un sistema integrato che non solo consente la diversità (tecnologica e disciplinare) ma anche la facilità di utilizzo permettendo agli scienziati di concentrarsi di più sulla loro ricerca lasciando le tecnologie d'interoperabilità a componenti e servizi esperti che non sono gestiti dalle comunità disciplinari.

Architettura del Sistema

Il sistema (vedi fig.1) prevede l'attivazione di alcuni nodi interconnessi e federati tra loro così denominati:

Common Node (CNR-DTA)

First level node (CNR, ENEA, INGV, OGS, MNA)

Second level node (Data Systems, banche dati...)

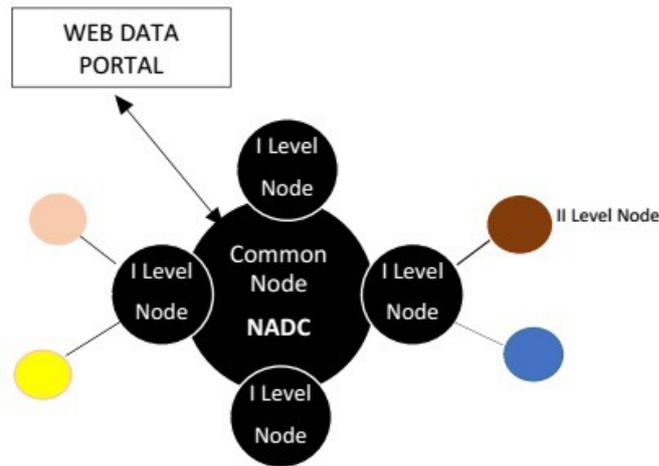


Fig.1 Schema logico dell'infrastruttura distribuita

Common Node

Rappresenta il nodo comune della infrastruttura dei dati e metadati polari nonché il National Antarctic Data Center (NADC) come da specifiche SCAR.

Il nodo comune è anche il portale Web per la ricerca e l'accesso ai dati e metadati che possono essere archiviati e conservati in data base distribuiti geograficamente ed appartenenti a Enti e Organizzazioni diverse.

Questo significa che il Common Node non è l'archivio centrale di tutti i dati e metadati polari, le informazioni infatti sono distribuite tra i vari nodi della infrastruttura, ed interconnessi tra loro attraverso servizi di mediazione (es. Brokering Approach).

First level Node

I nodi di primo livello sono sistemi eterogenei appartenenti ad Enti e Organizzazioni diverse che mettono in comune le loro capacità e conoscenza per dar vita ad un sistema più complesso. I nodi I livello devono stabilire un interoperability agreement con il Common node.

Second level Node

I nodi di II livello sono tipicamente dei data systems (o banche dati) nati per esigenze progettuali e/o sperimentali. Si interfacciano direttamente con un nodo di primo livello.

Il CNR è responsabile dello sviluppo della manutenzione e della operatività del Common Node e degli interoperability agreement verso i nodi di primo livello.

Il nodo di I livello è responsabile di stabilire gli interoperability agreement con il common node e di interfacciarsi con i nodi di livello II, III, ecc.

APPENDICE C

A titolo di esempio si riportano alcuni casi di licenze d'uso dei dati Creative Commons. Le licenze CC offrono [sei diverse articolazioni dei diritti d'autore](#) per artisti, giornalisti, docenti, istituzioni e, in genere, creatori che desiderino condividere in maniera ampia le proprie opere secondo il modello "alcuni diritti riservati". Il detentore dei diritti può non autorizzare a priori usi prevalentemente commerciali dell'opera (opzione Non commerciale, acronimo inglese: NC) o la creazione di opere derivate (Non opere derivate, acronimo: ND); e se sono possibili opere derivate, può imporre l'obbligo di rilasciarle con la stessa licenza dell'opera originaria (Condividi allo stesso modo, acronimo: SA, da "Share-Alike"). Le combinazioni di queste scelte generano le [sei licenze CC](#), disponibili anche in [versione italiana](#).

Simboli	Sigla	Descrizione
	CC BY	Permette di distribuire, modificare, creare opere derivate dall'originale, anche a scopi commerciali, a condizione che venga riconosciuta la paternità dell'opera all'autore.
	CC BY-SA	Permette di distribuire, modificare, creare opere derivate dall'originale, anche a scopi commerciali, a condizione che venga riconosciuta la paternità dell'opera all'autore e che alla nuova opera vengano attribuite le stesse licenze dell'originale (quindi ad ogni derivato verrà consentito l'uso commerciale). Questa licenza, per certi versi, può essere ricondotta alle licenze " copyleft " del software libero e open source .
	CC BY-ND	Permette di distribuire l'opera originale senza alcuna modifica, anche a scopi commerciali, a condizione che venga riconosciuta la paternità dell'opera all'autore.
	CC BY-NC	Permette di distribuire, modificare, creare opere derivate dall'originale, a condizione che venga riconosciuta la paternità dell'opera all'autore, ma non a scopi commerciali. Chi modifica l'opera originale non è tenuto ad utilizzare le stesse licenze per le opere derivate.
	CC BY-NC-SA	Permette di distribuire, modificare, creare opere derivate dall'originale, ma non a scopi commerciali, a condizione che venga riconosciuta la paternità dell'opera all'autore e che alla nuova opera vengano attribuite le stesse licenze dell'originale (quindi ad ogni derivato non sarà permesso l'uso commerciale).
	CC BY-NC-ND	Questa licenza è la più restrittiva: consente soltanto di scaricare e condividere i lavori originali a condizione che non vengano modificati né utilizzati a scopi commerciali, sempre attribuendo la paternità dell'opera all'autore.