

Linee guida per la creazione del Catalogo delle Banche Dati del CRA

1. Sommario

1.	Sommario.....	1
2.	Premessa	2
2.1.	Finalità del documento	2
2.2.	Glossario essenziale	2
3.	Catalogo delle Banche dati	3
3.1.	Informazioni presenti nel Catalogo	3
3.1.1.	Definizione di Banca dati	3
3.1.2.	Attributi	3
3.2.	Formato del Catalogo	5
3.2.1.	Titolo del dataset	5
3.2.2.	Campi dei metadati	5
3.2.3.	Campi dei dati	5
3.2.4.	Altre caratteristiche	6
4.	Censimento delle Banche dati	6
4.1.	Flusso informativo	6
4.2.	Creazione del Catalogo del Centro o Unità di Ricerca	7
4.2.1.	Nome del file	7
4.2.2.	Metadati del dataset	7
4.2.3.	Dati delle Banche dati	7
5.	Aggiornamento del Catalogo	7

2. Premessa

Il CRA si pone l'obiettivo di soddisfare i requisiti dell'art. 9 del DL n. 179/2012, convertito in Legge n. 221/2012, che stabilisce di pubblicare nella sezione Amministrazione trasparente del portale "il **catalogo dei dati, dei metadati, e delle relative banche dati ... ed i regolamenti che ne disciplinano l'esercizio della facoltà di accesso telematico e il riutilizzo ...**".

A tale scopo il CRA sta predisponendo il proprio catalogo delle Banche dati.

Questa iniziativa si colloca tra le finalità di una più ampia strategia descritta nel documento "Strategia per la valorizzazione del patrimonio informativo del CRA" approvato dal CdA nella seduta del 6 novembre 2013 e pubblicato sul portale Web del CRA nella sezione "Amministrazione trasparente – Altri contenuti – Accessibilità e Catalogo di dati, metadati e banche dati"; esso è reperibile al seguente indirizzo: <http://sito.entecra.it/portale/public/documenti/PSI/cra-strategia-valorizzazione-patrimonio-informativo.pdf>.

2.1. Finalità del documento

Il presente documento contiene le indicazioni per la creazione di un Catalogo delle Banche dati in uso presso le Direzioni, i Centri e le Unità di Ricerca del Consiglio per la Ricerca e la sperimentazione in Agricoltura.

Il Catalogo delle Banche dati verrà gestito presso la Direzione Centrale Attività Scientifiche che raccoglierà le informazioni provenienti da tutte le strutture di Ricerca.

Per raggiungere lo scopo di pubblicare un Catalogo delle Banche dati le cui informazioni siano omogenee ed attendibili, nel presente documento vengono dettagliate le indicazioni necessarie per armonizzare il processo di raccolta ed aggiornamento delle informazioni. In particolare saranno indicate:

- a. Le definizioni, informazioni e formati che costituiranno il Catalogo
- b. La descrizione del processo informativo per il censimento delle Banche dati
- c. La descrizione del processo informativo per l'aggiornamento del Catalogo

Il presente documento non contiene alcuna raccomandazione o prescrizione relativa al modo con cui le singole Banche dati vengono organizzate e gestite dalle strutture di Ricerca che sono quindi libere di adottare tecnologie, prodotti, metodi e strutture dati confacenti alle proprie esigenze informative ed operative.

2.2. Glossario essenziale

Al fine di evitare il rischio di fraintendimenti, è opportuno definire qui alcuni termini che rappresentano i principali concetti informatici così come vengono utilizzati all'interno di questo documento:

- Dati (*data*): informazioni rappresentate in un modo direttamente utilizzabile da un computer
- Metadati (*metadata*): dato-descrizione di un insieme di dati
- Informazione (*information*): conoscenza rappresentata in modo utilizzabile dall'uomo
- Sistema informatico: insieme di computer e programmi per computer adatti ad automatizzare processi informativi elaborando dati
- Sistema informativo: automazione di un processo informativo mediante l'uso di uno specifico Sistema informatico
- Applicazione *software*: insieme di programmi utilizzati all'interno di uno specifico sistema informativo
- Applicazione *Web*: applicazione software utilizzabile via Internet attraverso un browser standard
- Banca dati (*database*): insieme di dati ben organizzato per consentirne l'uso da parte di un'applicazione software
- *Dataset*: insieme di dati omogenei, eventualmente estratti da un database
- DBMS (*Database Management System*): applicazione software per gestire database e capace di rendere disponibili i dati ad un'altra applicazione software
- *Open data*: tipologie di dati liberamente accessibili a tutti, privi di brevetti o altre forme di controllo che ne limitino la riproduzione

- Dominio: corpus omogeneo delle conoscenze e delle opere d'ingegno attinenti un determinato campo del sapere. Nel presente contesto, la ricerca e la sperimentazione in agricoltura.
- “Esperto di dominio”: persona che possiede una conoscenza specifica di una disciplina, materia, ambito, settore della Scienza, capace di comprendere il significato dei dati inseriti all'interno di un *dataset*

3. Catalogo delle Banche dati

3.1. Informazioni presenti nel Catalogo

Ponendoci l'obiettivo di creare il Catalogo delle Banche dati del CRA, è indispensabile puntualizzare ciò che l'Ente intende esattamente con il termine “Banca dati”, ovvero quali proprietà devono o possono avere le entità di questa particolare classe di oggetti.

3.1.1. Definizione di Banca dati

Per gli scopi del presente documento, una Banca dati è un *database* o un *dataset* (un insieme di dati omogenei non presenti allo stesso modo su altro database) o un *paper* (un insieme di informazioni registrate su un supporto non digitale ma trasformabile in *dataset*) prodotto e/o mantenuto e/o pubblicato nell'ambito delle attività di ricerca e sperimentazione in agricoltura da una qualsiasi Direzione, Centro o Unità di ricerca del CRA.

Per gli scopi di questo documento, il Catalogo contiene oggetti a cui può essere associato il seguente significato:

Una Banca dati si identifica con l'informazione che essa contiene (**contenuto**), e non con il particolare strumento informatico di gestione del *database* o del *dataset* (**contenitore**). Tantomeno con la disponibilità di un'applicazione Web per l'accesso alle informazioni e ai dati.

Il contenuto informativo delle Banche dati del CRA costituisce quindi un “patrimonio” della ricerca e sperimentazione in agricoltura da valorizzare e salvaguardare, secondo le raccomandazioni della Commissione Europea in merito a “l'accesso aperto alle pubblicazioni scientifiche, ai dati della ricerca e alla conservazione e riutilizzo dell'informazione scientifica”¹.

Ogni singolo database o dataset, prodotto e/o mantenuto e/o pubblicato dal CRA, è quindi un'entità della classe Banca dati del CRA. Ogni entità è caratterizzata da diversi attributi i cui valori la rendono **unica**, **riconoscibile** ed **identificabile**.

3.1.2. Attributi

Nell'identificare gli attributi, si deve tener conto che essi sono di tre tipi.

Attributi obbligatori

Con attributo obbligatorio si indica quella tipologia d'informazione, concernente una banca dati, imposta o consigliata da normative e raccomandazioni governative in merito alla Trasparenza amministrativa o alla Valorizzazione del patrimonio informativo.

Le “Linee guida per la valorizzazione del patrimonio informativo pubblico (PSI - *Public Sector Information*)”² emanate dall'Agenzia per l'Italia Digitale raccomandano di indicare per ogni Banca dati le seguenti informazioni:

- **Titolare della banca dati**
“E' colui che all'interno dell'amministrazione è responsabile del procedimento amministrativo che popola la specifica

¹ Gazzetta ufficiale dell'Unione europea - L 194/39

Raccomandazione della Commissione del 17 luglio 2012 sull'accesso all'informazione scientifica e sulla sua conservazione
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:194:0039:0043:IT:PDF>

² Agenzia per l'Italia Digitale

Linee guida per la valorizzazione dei dati della PA

http://www.digitpa.gov.it/sites/default/files/allegati_tec/LG_Val_PSI_v1.0.pdf

fonte del dato, che ne cura la qualità e il relativo aggiornamento. E' tipicamente un Dirigente o un Quadro che coordina un gruppo di persone che svolgono il loro lavoro quotidiano attorno alla fonte del dato in esame. Ha anche il potere di decidere se modificare un certo dato sulla base di indicazioni pervenute ad esempio da cittadini che, vedendo il dataset, ne richiedono una versione evoluta."

- **Referente tecnico della banca dati**

"Si tratta tipicamente di un componente del gruppo coordinato dal Titolare della banca dati; esso ha conoscenze informatiche almeno di base e svolge un ruolo operativo sul sistema gestionale afferente al dato. Inoltre, fornisce indicazioni circa il reperimento concreto dei dati dalla base dati, e cura il monitoraggio dei vari "connettori" che a partire dalla base dati espongono il dato come Open Data. Tipicamente riceve materialmente le segnalazioni dei cittadini sul dataset di propria competenza, e le smista eventualmente al Referente tematico per valutarne il contenuto, prima di chiedere al Titolare della Banca Dati l'approvazione per eventuali azioni correttive strutturali sul dataset."

- **Referente tematico della banca dati**

"Si tratta di un esperto di dominio che conosce in modo approfondito l'ufficio e la storia dei dati su cui l'ufficio opera. Spesso propone nuovi dataset da esporre a partire dal sistema gestionale corrispondente e cura eventuali valutazioni di dominio o relative al significato dei dati. Ha anche la possibilità di compiere bonifiche e semplici adeguamenti sulla banca dati, su segnalazione di cittadini o su valutazioni proprie. Riferisce invece al Titolare della Banca dati la necessità di eventuali variazioni strutturali al sistema gestionale che insiste sui dati."

Nell'ambito della creazione del Catalogo delle Banche dati del CRA, le su citate raccomandazioni normative sono recepite.

Attributi necessari

Con attributi necessari si indica quella tipologia d'informazione ritenuta indispensabile dal CRA ai fini della costruzione del proprio Catalogo delle Banche dati.

Ai fini della redazione del Catalogo, oltre a quelli obbligatori, si considerano necessari i seguenti attributi:

- **Codice**
Si tratta di una stringa alfanumerica che identifica univocamente la Banca dati
- **Acronimo**
Si tratta di una sigla leggibile come singola parola o un nome breve con cui viene comunemente indicata la Banca dati
- **Nome**
Si tratta del nome per esteso della Banca dati
- **Tipo**
Si tratta della tipologia dell'insieme di dati (p.e. database, dataset, ecc.)
- **Descrizione**
Si tratta di una breve descrizione che identifica con precisione il contenuto informativo della Banca dati, mettendone in risalto il valore scientifico
- **URL**
Si tratta dell'indirizzo internet, se disponibile, da cui è possibile accedere via browser alle informazioni e ai dati della Banca dati
- **Proprietario**
Si tratta del Direzione, Centro o Unità di Ricerca del CRA che ha la responsabilità dei contenuti della Banca dati
- **Wikipedia**
Si tratta della URL di una eventuale pagina di Wikipedia dedicata alla Banca dati, in cui sono inserite le informazioni e i link alla Banca dati e alle principali risorse connesse

Attributi accessori

Ogni Banca dati è caratterizzata da innumerevoli proprietà di tipo scientifico, tecnico, amministrativo, ecc. Con attributi accessori si indica tutta l'informazione non direttamente utile allo scopo della creazione del Catalogo.

3.2. Formato del Catalogo

Il Catalogo delle Banche dati è un dataset pubblicato sul portale Web del CRA nella sezione "Amministrazione trasparente – Altri contenuti – Accessibilità e Catalogo di dati, metadati e banche dati" che presenta le seguenti caratteristiche:

- Contiene metadati che identificano il dataset
- Contiene dati che compongono il catalogo
- È realizzato nel formato elettronico OpenDocument (ODS)
- È pubblicato secondo una licenza che ne permette il libero uso, con l'obbligo di citare la fonte (IODL V2.0 - <http://www.dati.gov.it/iodl/2.0/>)

3.2.1. Titolo del dataset

CRA Databases List

3.2.2. Campi dei metadati

Le "Linee guida per la valorizzazione del patrimonio informativo pubblico (PSI – Public Sector Information)" precedentemente citato, raccomandano che ogni dataset pubblicato debba essere corredato da *"metadati obbligatori (ed eventualmente) metadati obbligatori condizionatamente al verificarsi di certe condizioni"*.

Per la creazione del Catalogo delle Banche dati, il relativo dataset conterrà i soli metadati obbligatori così raccomandati:

Campo	Provenienza	Descrizione e note
LICENSE	CC	Indica la licenza utilizzata
PUBLISHER	DC	Indica il soggetto che pubblica il dataset. Spesso coincide con creator
CREATOR	DC	Indica il soggetto che ha prodotto il dataset. Spesso coincide con publisher
TITLE	DC	Indica il titolo del dataset
DESCRIPTION	DC	Indica la descrizione in linguaggio naturale del dataset
MODIFIED	DCMI	Indica la data di ultimo aggiornamento
KEYWORD	DCAT	Indica le parole chiave, separate da virgole, che descrivono il dataset (tag)

Note

CC = Creative Commons
 DC = Dublin Core
 DCAT = Data Catalog Vocabulary
 DCMI = Dublin Core Metadata Initiative

Tali metadati sono inseriti in testa ai dati contenuti nel dataset.

Poiché è necessario dare la più ampia diffusione alle informazioni, sia le etichette che il contenuto dei campi sono previste essere in lingua inglese (EN).

3.2.3. Campi dei dati

Con riferimento alle proprietà precedentemente indicate, i campi indispensabili per inserire le informazioni riguardanti le Banche dati sono i seguenti.

Alcuni di questi campi contengono dati che devono essere inseriti allo stesso modo di come si presentano all'interno del dataset "CRA Directory", pubblicato sul portale Web del CRA ed azionabile dall'indirizzo:

<http://sito.entecra.it/portale/public/documenti/psi/cra-directory.ods>

Campo	Significato	Descrizione e note
CODE	Codice	Stringa alfanumerica che identifica univocamente la Banca dati (es. CRA-CER-000), formata da: <ol style="list-style-type: none"> 1. Sigla della Direzione, Centro o Unità di Ricerca, inserita così come codificata nel campo "Code" del dataset "CRA Directory" 2. Trattino (segno "meno") 3. Numero progressivo di tre cifre a partire da 000
ACRONYM	Acronimo	Sigla leggibile come singola parola o un nome breve con cui viene comunemente indicata la Banca dati (es. Identivit, Monitor, Agritrasfer, ecc.)
NAME	Nome	Nome per esteso della Banca dati
TYPE	Tipo	Tipologia dell'insieme di dati (database, dataset o paper)
DESCRIPTION	Descrizione	Breve descrizione in lingua inglese (EN) che identifica con precisione il contenuto informativo della Banca dati, mettendone in risalto il valore scientifico (max. 50 parole)
URL	URL	Indirizzo internet, se esistente, da cui è possibile accedere via browser alle informazioni e ai dati della Banca dati, nel seguente formato <i>http://nomehost.nomedominio/percorso</i> . In caso di database/dataset non accessibili via Internet, il campo viene riempito con l'informazione "localhost"
OWNER	Proprietario	Nome della Direzione, Centro o Unità di Ricerca del CRA che ha la responsabilità dei contenuti della Banca dati, inserito così come codificato nel campo "Name" del Dataset "CRA Directory"
CONTENT MANAGER	Referente tematico	Nome e cognome dell'esperto di dominio che conosce in modo approfondito il significato dei dati
TECHNICAL REPRESENTATIVE	Referente tecnico	Nome e cognome dell'esperto informatico che svolge un ruolo operativo sul sistema di gestione dei dati. In caso di dataset, il Technical representative può coincidere con il Content manager
RESPONSIBLE	Titolare	Nome e cognome del responsabile della struttura di ricerca che gestisce la specifica Banca dati
WIKIPEDIA	Wikipedia	URL di una eventuale pagina di Wikipedia dedicata alla Banca dati, in cui sono inserite le informazioni e i link alla Banca dati e alle principali risorse connesse

Le etichette che identificano le proprietà delle Banche dati sono in lingua inglese (EN), così come il contenuto di alcuni campi.

3.2.4. Altre caratteristiche

Nome del file

Il nome del file è `cra-databases-list.ods`

URL

Il file è azionabile dal portale web del CRA, al seguente indirizzo:

<http://sito.entecra.it/portale/public/documenti/psi/cra-databases-list.ods>

Content manager del catalogo

Il Servizio S1 – Programmazione scientifica e Relazioni internazionali della Direzione Centrale, attraverso un proprio referente tecnico (**creator**).

4. Censimento delle Banche dati

Il Catalogo viene gestito centralmente. La Direzione Scientifica si fa carico di inserire nel Catalogo i dati relativi sia alle Banche dati gestite dalla Direzione Scientifica, sia quelli provenienti dalle banche dati gestite dalle strutture di Ricerca.

4.1. Flusso informativo

Al fine di realizzare il censimento delle Banche dati, ogni struttura di Ricerca realizza il **proprio** Catalogo delle Banche dati da essa gestite, utilizzando i criteri e il modello sopra descritti.

I dati pervengono centralmente secondo un processo informativo non automatizzato, basato sull'invio di un file OpenDocument (ODS) via posta elettronica all'indirizzo open.data@entecra.it

4.2. Creazione del Catalogo del Centro o Unità di Ricerca

Per realizzare il Catalogo, ogni struttura periferica potrà impiegare come modello il file `cra-databases-list.ods` ed adattarlo semplicemente per lo scopo.

Sostanzialmente dovrà:

- a. Modificare il nome del file
- b. Inserire i metadati relativi alla struttura periferica
- c. Inserire i dati relativi alle proprie Banche dati

4.2.1. Nome del file

Il nome del file si differenzierà da quello utilizzato per il Catalogo generale `cra-databases-list.ods` poiché in luogo della sigla CRA, presenterà l'acronimo del Centro o Unità di Ricerca, così come codificato nel campo "Code" del dataset "CRA Directory".

Es. `cra-vit-databases-list.ods`

4.2.2. Metadati del dataset

Il dataset presenta i seguenti metadati.

TITLE	<i>sigla del centro o unità di ricerca</i> ⁱⁱ Databases List
DESCRIPTION	List of major databases for finding and accessing scientific data, articles, images, and other collections of datasets and documents
PUBLISHER	<i>centro o unità di ricerca</i> ⁱⁱⁱ (nome in lingua inglese ^{iv})
CREATOR	<i>nome e cognome</i>
MODIFIED	<i>aaaa/mm/gg</i>
LICENSE	IODL V2.0 - http://www.dati.gov.it/iodl/2.0/
KEYWORD	CRA, Scientific databases, Scientific datasets, Agriculture, Open Data

Note:

- i. i valori evidenziati in grassetto devono essere obbligatoriamente gli stessi
- ii. acronimo del Centro o Unità di Ricerca, così come codificato nel campo "code" del dataset "CRA Directory"
- iii. nome del Centro o Unità di Ricerca, così come codificato nel campo "name" del dataset "CRA Directory"
- iv. nome in lingua inglese del Centro o Unità di Ricerca, così come codificato nel campo "description" del dataset "CRA Directory"

4.2.3. Dati delle Banche dati

Per l'inserimento dei dati, valgono le linee guida precedentemente riportate al par. 3.2.3.

5. Aggiornamento del Catalogo

Al verificarsi di una qualsiasi variazione sui dati del Catalogo, la struttura di Ricerca dovrà inviare il file aggiornato all'indirizzo di posta elettronica open.data@entecra.it avendo cura di aggiornare il valore del campo "Modified" dei metadati.

Centralmente verrà effettuato il lavoro di storicizzazione, conservando le versioni precedenti del solo file `cra-databases-list.ods` (Catalogo generale).