

Lorenzo Bacci*

Alimentare e allineare il *Nuovo soggettario* con le voci di classificazione e indicizzazione della banca dati DoGi – Dottrina Giuridica

La banca dati DoGi

DoGi – Dottrina Giuridica è una banca dati di riferimenti bibliografici e abstract di articoli pubblicati nelle riviste giuridiche italiane. La banca dati, istituita nel 1970, è un prodotto delle attività di ricerca in tema di accesso e diffusione dell'informazione giuridica condotte dall'Istituto di Informatica Giuridica e Sistemi Giudiziari del Consiglio Nazionale delle Ricerche (IGSG-CNR¹), con sede a Firenze (dal 1968 al 2001 tale istituto era chiamato "IDG - Istituto per la documentazione giuridica", dal 2002 al 2019 "ITTIG").

Nell'arco della sua storia sono stati spogliati oltre mezzo milione di contributi provenienti da circa 800 riviste. Attualmente la banca dati DoGi viene aggiornata bimestralmente e si incrementa ogni anno con lo spoglio di circa 13.000 contributi relativi a 297 riviste, delle quali 221 in classe A e 56 ad accesso aperto.

Lo spoglio dei contributi consiste nell'inserimento di una serie di metadati e informazioni aggiuntive che vanno ad arricchire il mero riferimento bibliografico, come abstract e sommari dell'articolo (con supporto multi lingua), fonti legislative o giurisprudenziali significative o addirittura oggetto del contributo, come nel caso di una nota sentenza, la classificazione del contributo per area del diritto, la selezione di una o più parole chiave di dominio giuridico, etc.

Nel quinquennio 2018-2023 l'Istituto di Informatica Giuridica e Sistemi Giudiziari ha portato avanti un complesso progetto che ha comportato il totale rifacimento dei sistemi redazionali e di pubblicazione della banca dati DoGi. Nell'ambito di tale progetto è stato sviluppato *DoGi Web*², la nuova piattaforma di consultazione della banca dati DoGi.

DoGi Web (fig. 1) permette la ricerca, la navigazione e la visualizzazione non solo degli spogli dei contributi, ma anche di tutte le risorse a essi associate: fascicoli, riviste ed editori, autori e curatori, fonti legislative e giurisprudenziali, classificazioni e parole chiave.

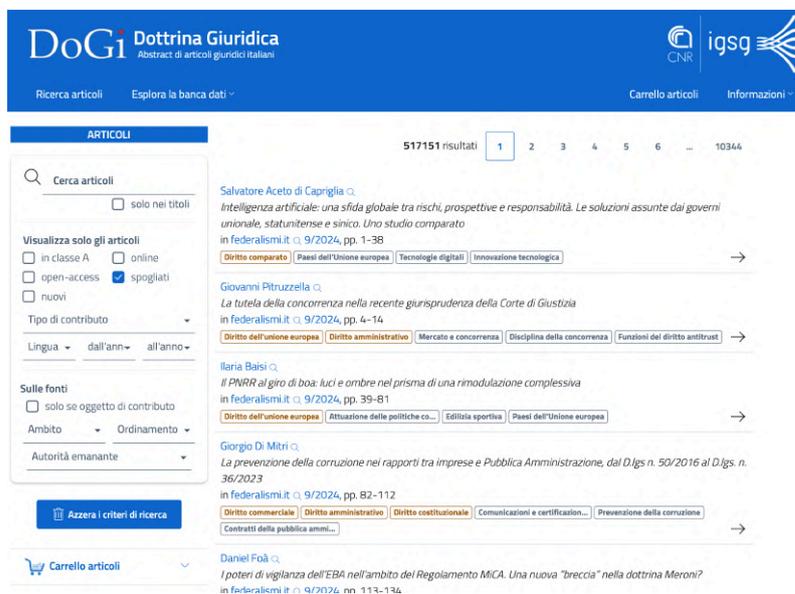


Fig. 1. Dogi Web, il sito di consultazione della banca dati DoGi – Dottrina Giuridica.

A differenza del precedente sito di consultazione della banca dati, *DoGi Web* segue una filosofia orientata al *semantic web* e ai *linked data*: ogni risorsa esposta dalla banca dati possiede infatti un proprio identificatore logico, ovvero una URI (*Uniform Resource Identifier*), attraverso il quale si può sia accedere alla risorsa via web sia farvi riferimento all'interno di una tripla RDF³.

Dove possibile, le URI sono anche “parlanti”, ovvero composte usando identificatori logici noti o informazioni che descrivono univocamente la risorsa stessa. Ad esempio, la URI della rivista “ADL Argomenti di diritto del lavoro” è composta dal *namespace* DoGi, un prefisso sempre presente, seguita dall'ISSN della rivista stessa: <http://dati.igsg.cnr.it/dogi/1126-5760>. Un fascicolo di tale rivista è identificato attraverso l'aggiunta del numero e dell'anno: http://dati.igsg.cnr.it/dogi/1126-5760/2024_2. Infine, per i contributi del fascicolo è possibile utilizzare i numeri di pagina: http://dati.igsg.cnr.it/dogi/1126-5760/2024_2/215-222.

Grazie allo sviluppo di *DoGi Web* e alla sua filosofia di esposizione dei dati che abbraccia, facilita e stimola i collegamenti con altri *dataset*, è stato possibile realizzare uno degli obiettivi principali previsti nell'ambito della collaborazione tra BNCF e IGSG, ovvero un incremento di interoperabilità tra le rispettive banche dati.

Il nuovo sistema di classificazione

Nell'ambito del rifacimento dei sistemi redazionali della banca dati DoGi sono stati parzialmente rivisti anche i sistemi di classificazione e indicizzazione dei contributi spogliati.

Attualmente sono disponibili:

- una tassonomia delle aree del diritto;
- un elenco di parole chiave significative per il dominio giuridico, trasversali alle aree del diritto, non gerarchicamente organizzate.
- Mentre le voci di classificazione DoGi che compongono la tassonomia delle aree del diritto possono essere considerate piuttosto costanti nel tempo, la lista delle parole chiave è soggetta a essere arricchita a ogni aggiornamento della banca dati, previa supervisione della redazione.
- In fase di spoglio è dunque possibile:
 - classificare un contributo in uno o più rami del diritto;
 - associare un contributo a una o più parole chiave (fig. 2).

Attualmente sono presenti nella banca dati DoGi 1571 voci di classificazione e 5369 parole chiave.

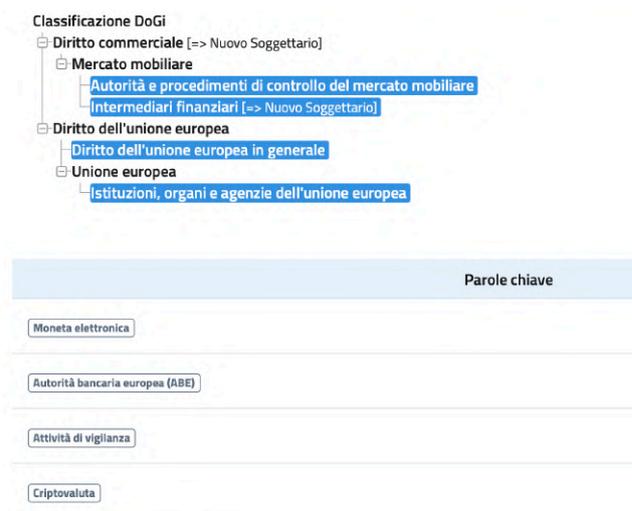


Fig. 2. Particolare delle voci di classificazione e delle parole chiave associate a un contributo in fase di spoglio e visualizzate su DoGi Web.

The screenshot shows the DoGi Web interface. At the top, there is a navigation bar with 'DoGi Dottrina Giuridica' and 'Abstract di articoli giuridici italiani'. Below this, there are search and navigation options. The main content area is titled 'Diritto d'autore' and displays a classification tree. The tree includes categories like 'Diritto Industriale', 'Concorrenza', and 'Diritto d'autore'. Under 'Diritto d'autore', there are sub-categories such as 'Atti di disposizione del diritto d'autore', 'Diritti connessi al diritto d'autore', 'Diritti dell'autore', 'Opere dell'ingegno protette', 'Principio di territorialità e convenzioni internazionali sul diritto d'autore', 'Requisiti per la tutela del diritto d'autore', 'Soggetti del diritto d'autore', 'Tutela del diritto d'autore', 'Utilizzazioni libere', 'Diritto industriale in generale', 'Privative industriali', and 'Segni distintivi'. Below the classification tree, there is a section titled 'Gli articoli più recenti classificati con questo termine:' which lists three articles with their authors and titles, each with a right-pointing arrow indicating a link to the article.

Fig. 3. La pagina relativa alla voce di classificazione Diritto d'autore su DoGi Web.

Integrazione Web

Grazie all'allineamento dei termini del *Nuovo soggettario* (NS) con le voci di classificazione e con le parole chiave della banca dati DoGi è stato possibile creare dei collegamenti ipertestuali per permettere la navigazione dal sito web del NS a *DoGi Web* e viceversa. In particolare, sul sito del NS, nel caso in cui un termine abbia un allineamento verso una voce di classificazione o una parola chiave in DoGi, viene visualizzato il link "DoGi" tra le fonti del termine.

Cliccando su tale *link* si accede a *DoGi Web*, nella pagina relativa alla corrispondente voce di classificazione o parola chiave, dove sono disponibili informazioni di contesto quali:

- il ramo completo nella tassonomia delle aree del diritto a cui la voce di classificazione afferisce e la possibilità di navigare tra le altre voci della tassonomia;
- i contributi più recenti presenti in banca dati che sono stati classificati con tale voce o associati a tale parola chiave;
- il numero complessivo di contributi spogliati e classificati o associati alla voce o alla parola chiave e la possibilità di visualizzarli tutti o di filtrarli ulteriormente attraverso le numerose opzioni di ricerca offerte da *DoGi Web*;
- un bottone che indica che la voce o la parola chiave ha un termine corrispondente nel NS, e attraverso cui è possibile compiere il percorso inverso, ovvero da *DoGi Web* alla scheda del termine allineato nel sito web del NS.

Ad esempio, nella scheda del termine del NS *Diritto d'autore*⁴ è presente tra le fonti il link che conduce alla pagina relativa alla voce di classificazione *Diritto d'autore*⁵ in *DoGi Web* (fig. 3).

Integrazione dati

Realizzare l'interoperabilità semantica tra *dataset* di banche dati eterogenee significa non solo individuare dei punti di contatto tra i *dataset*, nel nostro caso i termini allineati con le voci o le parole chiave, ma anche formalizzare i dati stessi secondo i costrutti e gli standard del *Semantic Web* e della pubblicazione di *linked data*.

Per questo motivo per la rappresentazione dei rispettivi lessici sia la BNCF sia l'IGSG hanno optato per l'impiego di SKOS.

SKOS⁶ (*Simple Knowledge Organization System*) è una raccomandazione del W3C (*World Wide Web Consortium*) ed è utilizzato per la rappresentazione di thesauri, lessici, glossari, schemi di classificazione, tassonomie, vocabolari controllati, ecc.

Alla base del formalismo di SKOS ci sono i *Concepts* che sono utilizzati nel nostro caso per rappresentare i singoli termini del NS e le singole voci di classificazione e parole chiave di DoGi. I *concepts* possono essere organizzati gerarchicamente usando le proprietà di SKOS *broader* e *narrower* e possono inoltre essere raccolti in *Concept Schemes*, che rappresentano i singoli vocabolari o lessici.

Da sottolineare infine che SKOS mette a disposizione delle proprietà specifiche per l'allineamento (o *mapping*) di *concepts* che appartengono a *concept schemes* diversi, come ad esempio la proprietà *closeMatch*.

Nell'immagine (fig. 4) vengono mostrati due *Concept Scheme*: il primo, prodotto da BNCF e relativo al NS, ha tra i suoi concetti di testa *Diritti assoluti* mentre il secondo, prodotto da IGSG e relativo alle voci di classificazione DoGi, ha *Diritto industriale*. Entrambi questi *top concepts* vedono tra i loro discendenti il concetto *Diritto d'autore*. Al fine di realizzare l'allineamento tra i due lessici, i due *concepts* sono dunque messi in relazione attraverso la proprietà SKOS *closeMatch*, che agisce in entrambe le direzioni (*symmetric property*).

Come detto, SKOS fa parte degli strumenti standard del Web Semantico, il quale ha come elemento fondante il linguaggio RDF (*Resource Description Framework*), un formalismo per esprimere conoscenza sotto forma di affermazioni soggetto-predicato-oggetto, in cui ogni elemento della *tripla* è identificato da una URI.

A sua volta RDF può essere scritto seguendo varie notazioni sintattiche. Uno dei più popolari formati è Turtle (*Terse RDF Triple Language*), che è stato utilizzato da IGSG per esportare e condividere i *dataset* SKOS relativi alle voci di classificazione e alle parole chiave della banca dati DoGi.

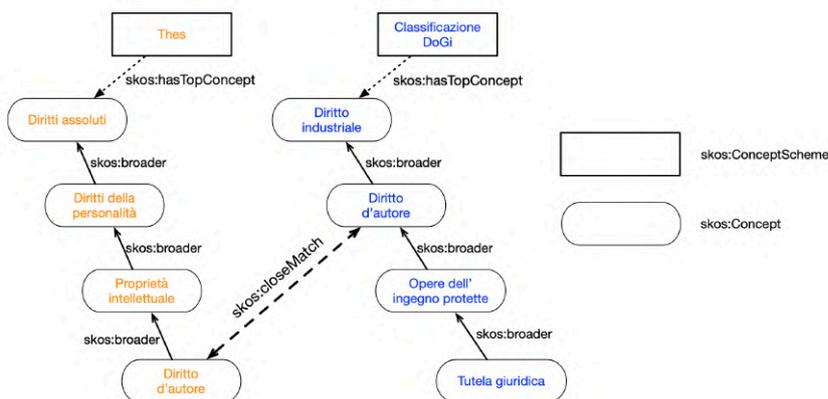


Fig. 4. Formalizzazione di lessici in SKOS e allineamento (mapping) dei concepts.

```

<http://dati.igsg.cnr.it/dogi/classifications/910>
  a                skos:Concept ;
  skos:inScheme    dogi:classificazioniDoGi ;
  skos:broader     <http://dati.igsg.cnr.it/dogi/classifications/1203> ;
  skos:narrower    <http://dati.igsg.cnr.it/dogi/classifications/1419> ,
<http://dati.igsg.cnr.it/dogi/classifications/1540> ,
<http://dati.igsg.cnr.it/dogi/classifications/1519> ,
<http://dati.igsg.cnr.it/dogi/classifications/1530> ,
<http://dati.igsg.cnr.it/dogi/classifications/1510> ,
<http://dati.igsg.cnr.it/dogi/classifications/1507> ,
<http://dati.igsg.cnr.it/dogi/classifications/1476> ,
<http://dati.igsg.cnr.it/dogi/classifications/1459> ,
<http://dati.igsg.cnr.it/dogi/classifications/1491> ;
  skos:closeMatch <http://purl.org/bn cf/tid/8541> ;
  dogi:bn cfLink  "https://thes.bn cf.firenze.sbn.it/termine.php?id=8541" ;
  rdfs:label      "Classificazione DoGi 910" ;
  skos:prefLabel  "DIRITTO D'AUTORE"@it .

```

Fig. 5. Estratto del dataset SKOS prodotto da IGSG in formato Turtle relativo alle voci di classificazione DoGi.

Nella prima riga di questo estratto (fig. 5) in formato Turtle del *dataset* relativo alle voci di classificazione DoGi è presente in verde la URI relativa al soggetto delle triple, in grassetto ci sono i predicati e alla loro destra sono presenti i valori, ovvero gli oggetti. Dunque vediamo che la risorsa in questione è uno SKOS *Concept* con *label* primaria “DIRITTO D’AUTORE”, che afferisce al *Concept Scheme*

classificazioniDoGi, ha uno *skos concept* più in alto nella gerarchia (*broader*) e vari *concepts* sotto di sé (*narrower*). In rosso vediamo le triple coinvolte nell’allineamento con il *dataset* BNCF relativo ai termini del NS: la proprietà DoGi *bn cfLink* è utilizzata per memorizzare il collegamento ipertestuale da *DoGi Web* alla scheda relativa al termine *Diritto d’autore* sul sito web del NS, mentre la proprietà SKOS *closeMatch* è utilizzata per l’allineamento e l’integrazione dei due lessici a livello dati.

Risultati e conclusioni

Nel primo tentativo di allineamento, avviato all’inizio della collaborazione tra BNCF e ITTIG, furono creati soltanto i collegamenti ipertestuali tra i rispettivi siti web in modo non simmetrico: 436 termini del NS avevano un link verso il precedente sito di consultazione della banca dati DoGi, mentre solo 217 voci del sistema di classificazione DoGi erano collegate al NS.

In seguito al rifacimento dei sistemi redazionali e di pubblicazione della banca dati DoGi e alla ripresa delle attività della collaborazione i termini del NS collegati con *DoGi Web* sono saliti a 813 (295 voci di classificazione e 518 parole chiave), e ognuno di questi presenta il percorso inverso, da *DoGi Web* al sito web del NS.

La collaborazione tra BNCF e IGSG, che aveva inizialmente permesso al NS di arricchirsi con l’aggiunta di numerosi termini di dominio giuridico, ha oggi raggiunto anche il suo secondo obiettivo, ovvero l’incremento dell’interoperabilità tra le rispettive banche dati. Questo è stato realizzato attraverso un’evoluzione dei sistemi redazionali e di pubblicazione della banca dati DoGi orientata al mondo dei *linked data* e grazie alla formalizzazione delle strutture dati relative al NS, alle voci di classificazione e alle parole chiave di DoGi seguendo gli standard e gli strumenti del Web Semantico come SKOS e RDF.

Entrambi i *dataset* SKOS relativi alle voci di classificazione e alle parole chiave prodotti da IGSG sono regolarmente aggiornati e disponibili per il download dal sito *DoGi Web*, garantendo così la possibilità di continuare ad arricchire e allineare il NS con le nuove voci di classificazione e in particolare con le nuove parole chiave della banca dati DoGi.

La collaborazione tra BNCF e IGSG è stata e continua a essere proficua, ha ottenuto risultati tangibili in termini di interoperabilità tra banche dati eterogenee e ha aperto la strada a nuove sfide e future opportunità, verso una sempre maggiore qualità e fruibilità delle risorse giuridiche.

NOTE

* Consiglio Nazionale delle Ricerche.

¹ <https://www.igsg.cnr.it/>

² <http://dati.igsg.cnr.it/dogi>

³ https://en.wikipedia.org/wiki/Semantic_triple

⁴ <https://thes.bnconfirenze.sbn.it/termine.php?id=8541>

⁵ <http://dati.igsg.cnr.it/dogi/classifications/910>

⁶ <https://www.w3.org/2004/02/skos/>

BIBLIOGRAFIA

Faro - Peruginelli 2017: S. Faro, G. Peruginelli (a cura di), *La dottrina giuridica e la sua diffusione*, Torino 2017, pp. 229-310.

Marinai - Peruginelli 2015: E. Marinai, G. Peruginelli, *La banca dati DoGi e la condivisione dei dati giuridici: nuovi orizzonti*, in O. Bonora, D. Coltellacci, L. Garbolino, M.C. Piazza, B. Paradiso, A. Perin, E. Secinaro (a cura di), *Ecosistemi per la ricerca Atti Convegno ACNP/NILDE*. Trieste, 22-23 maggio 2014, Trieste 2015, pp. 57-74.

Peruginelli 2005: G. Peruginelli, *The DoGi database: access to Italian legal literature*, in "Legal Information Management", vol. 5, n. 3, 2005, pp. 175-180.