



Requisiti di Qualità per Quality at Source (QaS)

Sommario

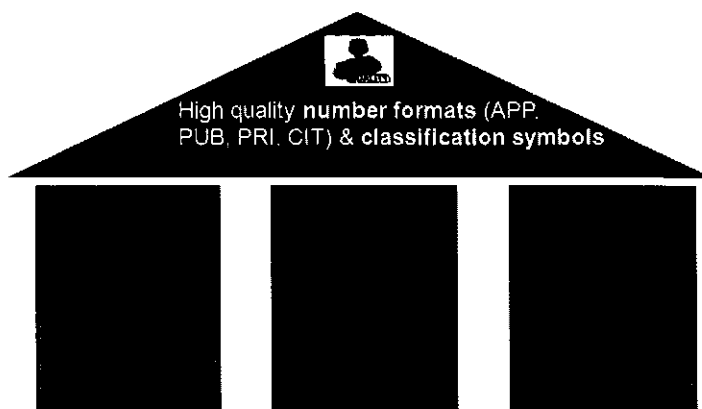
1.	Introduzione	2
2.	DATI BIBLIOGRAFICI STANDARDIZZATI	2
2.1.	CAMPI DI DATI BIBLIOGRAFICI MINIMI RICHIESTI (IN INGLESE)	2
2.2.	ALTRI CAMPI IMPORTANTI	2
2.3.	FORMATI NUMERICI E CODICI TIPO COERENTI (SECONDO GLI STANDARD WIPO 6, 10/C, 13, 14 E 16) PER.....	3
2.4.	DATE STANDARDIZZATE, CORRISPONDENTI AL LIVELLO DI PUBBLICAZIONE CORRETTO.....	3
2.5.	INFORMAZIONI DI CLASSIFICAZIONE STRUTTURATE SECONDO ST.8.....	3
3.	Formato DEI DATI.....	3
3.1.	ELETTRONICO, STRUTTURATO E CONFORME AGLI STANDARD.....	3
3.2.	Standard WIPO ST.36	4
3.3.	PDF PER I DATI IMMAGINI	5
4.	SCAMBIO DI DATI <i>Medialess</i>	5
4.1.	FTP	5
4.2.	SERVIZI WEB.....	5
5.	CORRETTEZZA, COMPLETEZZA E PUNTUALITÀ	6
5.1.	CORRETTEZZA	6
5.2.	COMPLETEZZA	6
5.3.	PUNTUALITÀ.....	7
5.4.	QUALITÀ DEI DATI	7
6.	RIEPILOGO	7
7.	ALLEGATO 1: GUIDA DI CONSULTAZIONE RAPIDA QaS	8

1. INTRODUZIONE

Il concetto Quality at Source (QaS) è stato elaborato per invitare gli Uffici IP a svolgere un ruolo attivo ai fini della generazione di dati di elevata qualità.

È costituito da quattro blocchi strutturali:

1. Dati bibliografici standardizzati:
 - a. Campi di dati minimi
 - b. Numeri e date strutturate
2. Formato dei dati (WIPO ST.36¹ e PDF)
3. Scambi *medialess*
4. Correttezza, completezza e puntualità



2. DATI BIBLIOGRAFICI STANDARDIZZATI

2.1. CAMPI DI DATI BIBLIOGRAFICI MINIMI RICHIESTI (IN INGLESE)

- Informazioni relative alla richiesta/presentazione (numero, data e tipo)
- Informazioni di priorità (numero/i, data/e e tipo di richiesta/e)
- Informazioni relative alla pubblicazione (numero e data pubblicazione/registrazione/brevetto)
- Tipo di pubblicazione / Codice tipo
- International Patent Classification (IPC) o Cooperative Patent Classification (CPC), laddove pertinenti
- Documenti correlati (es. richieste PCT, richieste divisionali, emendamenti)
- Titolo dell'invenzione (per le richieste regionali/PCT)
- Riferimenti citati
- Nome del richiedente (traslitterazione in caratteri latini)
- Nome dell'inventore (traslitterazione in caratteri latini)

2.2. ALTRI CAMPI IMPORTANTI

- Sintesi in inglese (per richieste regionali/PCT)
- Dati di classificazione nazionale
- Titolo dell'invenzione (lingua originale)
- Paese di residenza del richiedente
- Paese di residenza dell'inventore

¹ ST.36: [Processing of patent information using XML \(November 2007\)](#)

3.2. STANDARD WIPO ST.36

Lo standard per il formato dei dati sarà lo Standard WIPO ST.36¹¹, utilizzato per:

- Dati bibliografici (comprese la sintesi e le informazioni di classificazione)
Per i titoli e le sintesi, i codici lingua dovranno essere basati sulla norma ISO 639-1¹².
Idealmente, le informazioni relative ai nomi del richiedente e dell'inventore dovranno essere fornite con nome e cognome differenziati.
- Citazioni
I dati delle citazioni dovranno essere conformi a WIPO ST.14 e contenere:
 - riferimenti citati (letteratura brevettuale e non, in ogni caso strutturata¹³)
 - categorie (es. X, Y) e relative informazioni di rivendicazione
 - passaggi (es. pagina 2, righe 12-17, paragrafo/i, colonna/e)
 - origine del riferimento (es. PCT, nazionale, UIP, R141, ecc.)
 - fase della procedura (es. Richiedente, Ricerca, Esame, Opposizione)
 - data in cui è stato citato il riferimento
- Dati full-text
I dati full-text dovranno contenere informazioni di indicizzazione relative alle sottosezioni di un documento brevettuale, conformemente a WIPO ST.36, con almeno:
 - la descrizione
 - le rivendicazioniLe figure e le formule integrate dovranno essere fornite sotto forma di immagini TIFF (conformi a WIPO ST.36)
Come raccomandato in WIPO ST.36, tutti i caratteri dovranno essere codificati in Unicode UTF-8.
Per i dati full-text generati tramite un processo OCR, si applicherà un tasso minimo di precisione OCR del 98% (un tasso di precisione del 99,5% si applicherà invece per la qualità di traduzione).

Le caratteristiche di ST.36 sono elencate qui di seguito:

- Raccomandazione internazionale per l'elaborazione delle informazioni brevettuali tramite XML
- Nessun singolo modello di tag validi e dei loro rapporti, solo una raccomandazione (il "xx-documento-brevettuale"¹⁴ DTD)
- Definisce una serie di elementi XML, detti International Common Elements (ICE), che possono essere liberamente combinati da ogni Ufficio
- EPO raccomanda l'utilizzo del "semplice-documento-brevettuale" DTD¹⁵, uno schema XML integralmente conforme a ST.36, ma meno complesso, che prende in esame i tag più essenziali

¹¹ ST.36: Processing of patent information using XML (November 2007)

¹² ISO 639-1: Codes for the Representation of Names of Languages

¹³ In questo contesto, "strutturato" significa che ogni elemento è contrassegnato da un tag XML distinto, cioè senza una singola stringa free-text.

¹⁴ Cf. WIPO ST.36 Supplementary Material: http://www.wipo.int/standards/en/xml_material/st36/

¹⁵ Le più recenti definizioni del "semplice-documento-brevettuale" DTD possono essere acquisite tramite EPO Front Office (frontoffice@epo.org)

2.3. FORMATI NUMERICI E CODICI TIPO COERENTI (SECONDO GLI STANDARD WIPO² 6³, 10/C⁴, 13⁵, 14⁶ E 16⁷) PER

- Richiesta/presentazione (WIPO ST.13)
- Priorità (WIPO ST. 10/C)
- Pubblicazione/registrazione (WIPO ST.6 e ST.16)
- Citazione (WIPO ST.14)

2.4. DATE STANDARDIZZATE, CORRISPONDENTI AL LIVELLO DI PUBBLICAZIONE CORRETTO

Le date di calendario devono essere leggibili elettronicamente per consentire la corretta elaborazione automatizzata dei dati. Devono essere riportate tramite i numeri arabi del giorno, del mese e dell'anno, nelle sequenze descritte nello standard WIPO ST.2⁸.

La data di pubblicazione è uno degli elementi più essenziali delle informazioni di brevetto, in quanto indica la "data di messa a disposizione del pubblico", un concetto definito nello Standard WIPO ST.9⁹ (4.f). Deve sempre essere in linea con il giusto livello di pubblicazione, come riportato in INID 40, WIPO ST.9, dove sono specificate otto diverse date di pubblicazione.

2.5. INFORMAZIONI DI CLASSIFICAZIONE STRUTTURATE SECONDO ST.8

Per fare sì che la presentazione e lo scambio dei simboli di classificazione IPC siano leggibili elettronicamente ed elaborati correttamente, tali simboli devono rispettare una definizione di campo a lunghezza fissa, come riportato nello Standard WIPO ST.8¹⁰.

3. FORMATO DEI DATI

3.1. ELETTRONICO, STRUTTURATO E CONFORME AGLI STANDARD

I principali requisiti del formato dei dati sono i seguenti:

- Disponibilità elettronica
- Strutturato
- Conforme agli standard

Questi requisiti consentiranno l'elaborazione, la verifica e il caricamento automatizzati dei dati forniti. È quindi essenziale che gli uffici IP coinvolti vi si adeguino strettamente.

² Standard WIPO: <http://www.wipo.int/standards/en/>

³ ST.6: [Numbering of published patent documents \(December 2002\)](#)

⁴ ST.10/C: [Presentation of bibliographic data components \(March 2011\)](#)

⁵ ST.13: [Numbering of applications for IPRs \(February 2008\)](#)

⁶ ST.14: [References cited in patent documents \(February 2008\)](#)

⁷ ST.16: [Identification of different kinds of patent documents \(May 1997\)](#)

⁸ ST.2: [Manner for designating calendar dates \(May 1997\)](#)

⁹ ST.9: [Bibliographic data on and relating to patents and SPCs \(June 2013\)](#)

¹⁰ ST.8: [IPC symbols on machine-readable records \(March 2011\)](#)

3.3. PDF PER I DATI IMMAGINI

- Conformità PDF 1.3¹⁶ o versione successiva¹⁷, segnalibri con sottosezioni del brevetto e struttura ad albero di primo livello.
- Le informazioni di indicizzazione relative alle sottosezioni di un documento brevettuale dovranno essere conformi allo standard WIPO ST.33¹⁸ e includere i seguenti elementi (se disponibili):
 - Dati bibliografici
 - Sintesi
 - Descrizione
 - Rivendicazioni
 - Disegni
 - Emendamento
 - Report di ricerca

4. SCAMBIO DI DATI *MEDIALESS*

Lo scambio di dati *medialess* può avvenire in diversi modi. Quelli più comunemente utilizzati dall'EPO sono FTP e Servizi Web.

4.1. FTP

La maggior parte degli scambi di dati dell'EPO con Stati membri e non-membri avviene attualmente tramite un server FTP ospitato e gestito presso questo stesso Ente. Ciò consente il monitoraggio periodico dei flussi di dati in ingresso e un'elaborazione rapida dopo la loro convalida.

Per utilizzare il metodo FTP, le parti devono preventivamente concordare una **struttura di directory/cartelle** e una **convenzione per la denominazione dei file**¹⁹.

4.2. SERVIZI WEB

L'utilizzo di Servizi Web (Restful, Soap...) rappresenta la soluzione preferita. È possibile creare un URL per recuperare dati in un formato specifico, generato "on the fly" da un sistema database di backend. Esistono due possibilità per implementare i Servizi Web: hosting da parte dell'EPO per raccogliere i dati oppure da parte degli uffici nazionali per fornire i dati.

¹⁶ PDF 1.3:

http://www.images.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference13.pdf

¹⁷ Cf. Adobe PDF Reference Archives: http://www.adobe.com/devnet/pdf/pdf_reference_archive.html

¹⁸ ST.33: Data exchange of facsimile information of patent documents (February 1999)

¹⁹ Il set EPO per la struttura di directory/cartelle preferita e la convenzione per la denominazione dei file possono essere acquisiti tramite EPO Front Office (frontoffice@epo.org)

L'implementazione dei Servizi Web dovrà comprendere come minimo le quattro funzionalità seguenti:

1. In seguito ad una richiesta di informazioni, ottenimento di un elenco di numeri di pubblicazioni entro un determinato intervallo temporale
2. Possibilità di scaricare a richiesta i dati bibliografici di ciascuna di queste pubblicazioni in ST.36
3. Possibilità di scaricare a richiesta l'immagine di ciascuna di queste pubblicazioni in PDF con segnalibri
4. Informazioni sulla normale data/ora in cui vengono messi a disposizione nuovi dati, per consentire l'automazione

5. CORRETTEZZA, COMPLETEZZA E PUNTUALITÀ

5.1. CORRETTEZZA

Si tratta della precisione del formato e dei contenuti dei dati:

- XML correttamente formato e conforme agli standard, convalidato rispetto a un DTD o XSD precedentemente distribuito
- Formati numerici e codici tipo armonizzati ben documentati prima dello scambio per poter generare regole operative per la convalida automatica dei dati. Laddove possibile, i numeri di citazione, richiesta e priorità devono essere forniti in formato DOCDB e convalidabili tramite OPS Rest Services (Number Service²⁰)
- In alternativa, i numeri di richiesta/priorità possono essere inoltrati in WIPO ST.13
- I numeri di brevetto utilizzati nelle citazioni devono essere forniti come minimo nella forma: Codice-Paese (CC) seguito da un numero (numero variabile di cifre) e, se disponibile, da un codice tipo (es. A, A3, B1)
- Laddove possibile, le citazioni NPL devono essere fornite in forma strutturata o, in alternativa, sotto forma di stringa con virgola separatrice
- Soglia massima del 5% per i difetti dei dati nell'ambito di una determinata consegna

5.2. COMPLETEZZA

Fa riferimento all'intera copertura frontfile/backfile e al pacchetto dati completo:

- Disponibilità di tutti i tipi di dati per la stessa pubblicazione: dati bibliografici, immagini, citazioni e full-text
- Per i dati bibliografici e le immagini, deve esistere idealmente una corrispondenza del 100% con il **file di autorità**²¹ di tutte le pubblicazioni emesse fin dall'inizio. Gli uffici nazionali hanno il compito di fornire un file di autorità completo, in base alla specifica
- Le citazioni e i dati full-text devono essere completi a partire dal 1973.

Informazioni pertinenti facoltative

²⁰ OPS Number Service: <http://ops.epo.org/rest-services/number-service>

²¹ Le più recenti specifiche QaS Authority File, file XML campione AF DTD e AF possono essere acquisiti tramite EPO Front Office (frontoffice@epo.org)

- **File di contenuto dei dati forniti** per ogni consegna
Il file di contenuto dei dati forniti indica il numero totale di documenti inclusi in ogni nuova consegna e il relativo elenco dei numeri di pubblicazione. Si tratta di aggiornamenti periodici del file di autorità storico, se il formato è conforme alla specifica di quest'ultimo (cf. nota 21).

5.3. PUNTUALITÀ

Viene soddisfatta quando vi sono consegne periodiche e puntuali, alle date e nelle ore predefinite, tali da consentire la raccolta e il caricamento automatizzati dei dati.

Le consegne devono avvenire il più presto possibile dopo la data di pubblicazione (o preferibilmente prima).

5.4. QUALITÀ DEI DATI

L'Ufficio Brevetti eserciterà una ragionevole cura per garantire la precisione e la completezza dei suoi dati. Qualora ritenga che determinati dati debbano essere corretti, modificati o integrati, comunicherà tempestivamente tali correzioni, modifiche o integrazioni all'EPO, possibilmente a stretto giro di posta.

Alla loro ricezione, tutti i nuovi dati saranno sottoposti a test di compatibilità e conformità ai sistemi EPO. Una volta approvati dall'EPO, saranno caricati in base ai criteri e ai principi stabiliti dall'EPO. Qualora l'EPO rilevi che i dati forniti dall'Ufficio Brevetti sono difettosi, questi saranno rispediti all'Ufficio in questione per la loro verifica e, se necessaria, immediata correzione.

Tutte le modifiche apportate alle specifiche e alle caratteristiche dei dati dovranno essere notificate all'EPO²² almeno sei mesi prima della loro implementazione. Tale notifica dovrà includere una specifica completa dei dati, accompagnata da un file di prova rappresentativo.

6. RIEPILOGO

Al fine di conformarsi ai requisiti QaS, gli Uffici IP dovranno mettere a disposizione dei loro partner di scambio, in maniera puntuale, completa e corretta, almeno i seguenti cinque elementi di dati:

1. Dati bibliografici del brevetto in WIPO ST.36, contenenti tutti i campi minimi
2. Facsimile del documento (dati immagini) in PDF con segnalibri
3. Dati delle citazioni in WIPO ST.36
4. Dati full-text in WIPO ST.36
5. Trasferimento elettronico *medialess* di tutti i dati sopra indicati

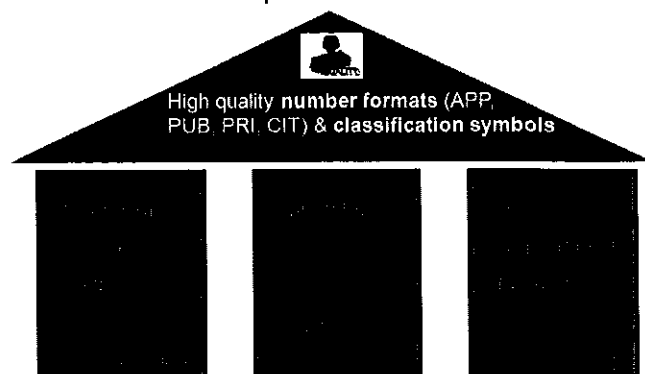
²² Le notifiche delle variazioni devono essere inviate a frontoffice@epo.org

7. ALLEGATO 1: GUIDA DI CONSULTAZIONE RAPIDA QAS

Il concetto Quality at Source (QaS) è stato elaborato per invitare gli Uffici IP a svolgere un ruolo attivo ai fini della consegna di dati brevettuali nazionali di elevata qualità.

È costituito da quattro blocchi strutturali:

1. Dati bibliografici standardizzati:
 - a. Campi di dati minimi
 - b. Numeri e date strutturate
2. Formato dei dati (WIPO ST.36 e PDF)
3. Scambi *medialess*
4. Puntualità, completezza e correttezza



1. Dati bibliografici standardizzati

- a. Campi di dati minimi e numeri/date strutturati
- b. Formati numerici e codici tipo coerenti, basati sugli standard WIPO, per i dati di richiesta/presentazione, priorità, pubblicazione/registrazione e citazione
- c. Date standardizzate, corrispondenti al livello di pubblicazione corretto
- d. Informazioni di classificazione strutturate in base agli standard WIPO (ST8)

2. Formato dei dati

- a. Elettronico, strutturato e conforme agli standard.
- b. **Standard WIPO ST.36** per i **dati bibliografici** (comprese sintesi e classificazione), **citazioni** (arricchite) e **full-text** indicizzato con immagini/formule integrate.
- c. **PDF 1.3** o versione successiva, **segnalibri con indicizzazione** delle sottosezioni del brevetto (ST.33) e struttura ad albero di primo livello.

3. Scambio di dati *medialess*

- a. **FTP** con struttura directory/cartelle e convenzione per la denominazione dei file preventivamente concordate.
- b. **Servizi Web** per la raccolta automatica dei dati.

4. Correttezza, completezza e puntualità

- a. **XML correttamente formato e conforme agli standard**, associato al DTD specificato.
- b. **Formati numerici e codici tipo armonizzati** e ben documentati.
- c. **Soglia massima del 5%** per i difetti dei dati nell'ambito di una determinata consegna.
- d. **La correzione, la modifica o l'integrazione dei dati** devono essere implementate e comunicate all'EPO (frontoffice@epo.org) **tempestivamente**.
- e. **Corrispondenza del 100% con il file di autorità** di tutte le pubblicazioni emesse per i dati bibliografici e le immagini (citazioni e full-text a partire dal 1973).
- f. **Disponibilità di tutti i tipi di dati** (es. dati bibliografici, immagini, citazioni e full-text) per la stessa pubblicazione.