



Requisiti di qualità per Progetto qualità dati (Qualità alla fonte)

Indice

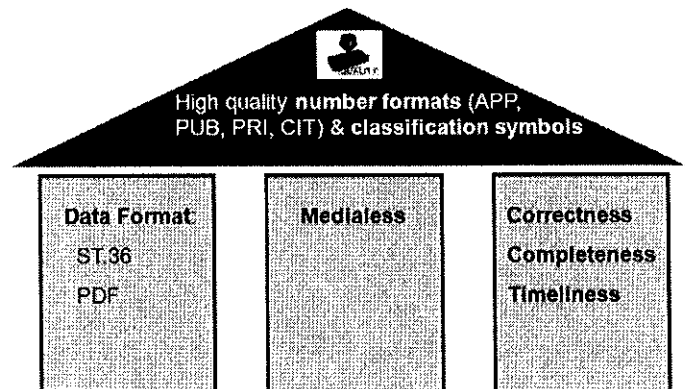
1. Introduzione	2
2. Dati bibliografici standardizzati	2
2.1. Campi dati bibliografici minimi obbligatori (in inglese)	2
2.2. Altri campi rilevanti	2
2.3. Formati numerici coerenti e codici tipo (basati sugli standard WIPO 6, 10/C, 13, 14 e 16)	2
2.4. Date standardizzate, corrispondenti al livello di pubblicazione corretto	3
2.5. Informazioni di classificazione strutturate secondo ST.8	3
3. Formato dati	3
3.1. Digitale, strutturato e conforme alle normative	3
3.2. Normativa WIPO ST.36	3
3.3. PDF per dati immagine	4
4. Consegna senza supporto fisico	7
4.1. FTP	7
4.2. Web Services	7
5. Correttezza, completezza, tempestività	7
5.1. Correttezza	7
5.2. Completezza	8
5.3. Tempestività	8
5.4. Qualità dati	8
6. Sintesi	9
7. Allegato 1: Guida di riferimento rapido QaS	10

Introduzione

Il concetto di Qualità alla fonte (QaS, Quality at Source) è stato creato per invitare gli uffici IP a svolgere un ruolo attivo nel garantire la produzione di dati di alta qualità.

Si basa su quattro elementi fondamentali:

1. Dati bibliografici standardizzati:
 - a. Campi dati minimi
 - b. Numeri e date strutturati
2. Formato dei dati (WIPO ST.36¹ e PDF)
3. Scambio senza supporto fisico
4. Corretto, completo e tempestivo



Dati bibliografici standardizzati

Campi dati bibliografici minimi obbligatori (in inglese)

- Informazioni sulla domanda/deposito (numero, data e tipo)
- Informazioni sulla priorità (numero, data e tipo)
- Informazioni sulla pubblicazione (pubblicazione/registrazione/numero e data del brevetto)
- Tipo di pubblicazione/Codice tipo
- Classificazione internazionale dei brevetti (IPC, International Patent Classification) o Classificazione dei brevetti cooperativi (CPC, Cooperative Patent Classification), ove applicabile
- Documenti correlati (es. domande PCT, domande divisionali, modifiche)
- Titolo dell'invenzione (per domande regionali/PCT)
- Riferimenti citati
- Nome del richiedente (traslitterazione latina)
- Nome dell'inventore (traslitterazione latina)

Altri campi rilevanti

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

Formati numerici coerenti e codici tipo (basati sugli standard WIPO² 6³, 10/C⁴, 13⁵, 14⁶ e 16⁷) per

- Domanda/deposito (WIPO ST.13)
- Priorità (WIPO ST. 10/C)
- Pubblicazione/registrazione (WIPO ST.6 e ST.16)

¹ ST.36: Processing of patent information using XML (November 2007)

² Normative 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)

- Citazione (WIPO ST.14)

Date standardizzate, corrispondenti al livello di pubblicazione corretto.

Le date del calendario devono essere leggibili dalla macchina per consentire una corretta elaborazione automatizzata dei dati. Dovrebbero essere indicate dai numeri arabi del giorno, del mese e dell'anno nelle sequenze descritte nella normativa WIPO ST.2⁸.

La data di pubblicazione è uno degli elementi più cruciali delle informazioni sui brevetti, poiché indica la "data di messa a disposizione al pubblico", un concetto definito nella normativa WIPO ST.9⁹ (4.f). Questa dovrebbe essere sempre in linea con il livello di pubblicazione appropriato, come indicato in INID 40, in WIPO ST.9, dove sono specificate otto diverse date di pubblicazione.

Informazioni di classificazione strutturate secondo ST.8

Per garantire che la presentazione e lo scambio dei simboli di classificazione IPC siano leggibili dalla macchina ed elaborati correttamente, i simboli di classificazione devono seguire una definizione di campo a lunghezza fissa come delineato nella normativa WIPO ST.8¹⁰.

Formato dati

Digitale, strutturato e conforme alle normative

I requisiti fondamentali per il formato dati sono

- Disponibilità digitale
- Strutturato
- Conforme alle normative

Questi requisiti consentiranno l'elaborazione, la verifica e il caricamento automatizzati dei dati forniti. È quindi fondamentale che gli uffici IP coinvolti seguano rigorosamente i requisiti.

Normativa WIPO ST.36

Lo standard del formato dei dati dovrebbe essere la normativa WIPO ST.36¹¹ e sarà utilizzato per:

- Dati bibliografici (compreso riassunto e informazioni sulla classificazione)
Per i titoli e i riassunti, i codici delle lingue dovrebbero essere basati sulla ISO 639-1¹²
Idealmente verranno fornite le informazioni sui nomi dei richiedenti e degli inventori nomi e cognomi differenziati (nome e cognome)
- Citazioni
I dati delle citazioni devono essere conformi a WIPO ST.14. Dovrebbero contenere:
 - riferimenti citati (letteratura brevettuale e non brevettuale, entrambe strutturate¹³)
 - categorie (ad es. X, Y) e le relative informazioni sulla rivendicazione

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)

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

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

¹³ Strutturato in questo contesto significa che ogni elemento è contrassegnato con un tag XML separato, quindi nessuna singola stringa di testo libero

- passaggi (es. pagina 2 righe 12-17, paragrafi colonne)
- 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 dovrebbero contenere informazioni di indicizzazione che coprano le sottoparti di un documento brevettuale secondo WIPO ST.36, riguardanti come minimo:

- Descrizione
- Rivendicazioni

Le immagini e le formule incorporate devono essere fornite come immagini TIFF (conforme a WIPO ST.36)

Come consigliato in WIPO ST.36, tutti i caratteri dovrebbero essere codificati in Unicode UTF-8

Per i dati full-text prodotti dal processo OCR, dovrebbe essere applicato un tasso di accuratezza OCR minimo del 98% (per la qualità della traduzione dovrebbe applicarsi un tasso di accuratezza superiore del 99,5%)

Le caratteristiche di ST.36 sono elencate di seguito:

- Raccomandazione internazionale per l'elaborazione delle informazioni sui brevetti mediante XML
- Nessun singolo modello di tag validi e relativa relazione, solo una raccomandazione (il DTD "documento-brevetto-xx"¹⁴)
- Definisce una serie di elementi XML denominati "International Common Elements" (ICEs), che possono essere liberamente combinati da qualsiasi ufficio
- L'EPO raccomanda di utilizzare il DTD "documento di brevetto semplice" fornito¹⁵, uno schema XML completamente conforme a ST.36 ma con una complessità ridotta che copre i tag più essenziali

PDF per dati immagine (Facsimile)

1. Formato PDF

Il PDF ricevuto dovrebbe essere:

- Conforme a PDF 1.7 o superiore.
- Aggiunto ai segnalibri (con le sottoparti del brevetto: bibliografia, descrizione, rivendicazioni, disegni, ecc.), i segnalibri devono essere una radice di primo livello.
- Linearizzato o non linearizzato.
- Né crittografato né protetto da password.

2. Segnalibri

¹⁴ Vedere materiale supplementare WIPO ST.36: http://www.wipo.int/standards/en/xml_material/st36/

¹⁵ Le ultime definizioni del DTD "documento di brevetto semplice" possono essere acquisite tramite EPO Front Office (frontoffice@epo.org)

I segnalibri sono obbligatori e devono essere nella lingua nazionale del brevetto o in alternativa in inglese, francese o tedesco.

Le 9 possibili sottoparti di un documento di brevetto sono:

- Dati bibliografici
- Rivendicazioni
- Disegni
- Modifiche
- Descrizione
- Riassunto
- Rapporto di ricerca
- Citazioni richiedente
- Sequenza DNA

Nel caso in cui i segnalibri siano consegnati in una lingua nazionale, è auspicabile ricevere dall'Ufficio nazionale un elenco delle 9 sottoparti PDF tradotte in quella lingua. Si trova di seguito un esempio di questo, ricevuto dall'Ufficio brevetti portoghese:

Sottoparte PDF Segnalibri PDF

Dati Biblio	Biblio
Rivendicazioni	Reivindicações
Disegni	Desenhos
Modifiche	Adições
Descrizione	Descrição
Riassunto	Resumo
Rapporto di ricerca	Search report
Citazioni richiedente	Applicant citations
Sequenza DNA	DNE Sequence

Inoltre, questi sono alcuni esempi di documenti correttamente contrassegnati ricevuti dall'EPO:

Esempio dalla Spagna (in spagnolo):

Primera Página
 Descripción
 Reivindicaciones
 Dibujos

OFICINA ESPAÑOLA DE PATENTES Y MARCAS
 ESPAÑA

2 198 509

TRADUCCIÓN DE PATENTE EUROPEA

Número de solicitud europea: 97402088.7
 Fecha de presentación: 10.02.1997
 Número de presentación de la solicitud: 0789479
 Fecha de publicación de la solicitud: 13.08.1997

Esempio dalla Corea (in coreano):

(19)대한민국특허청(KR)
 (12) 공개특허공보(A)

51 Int. CL7
 C01B 31/02
 B82B 3/00

(11) 공개번호 10-2004-0076884
 (43) 공개일자 2004년09월03일

Esempio dalla Svizzera (in inglese):

Abstract
 Description
 Claims
 Drawings

SCHWEIZERISCHE EidGENOSSENSCHAFT
 EidGENÖSSISCHES Institut für GEISTIGES EIGENTUM

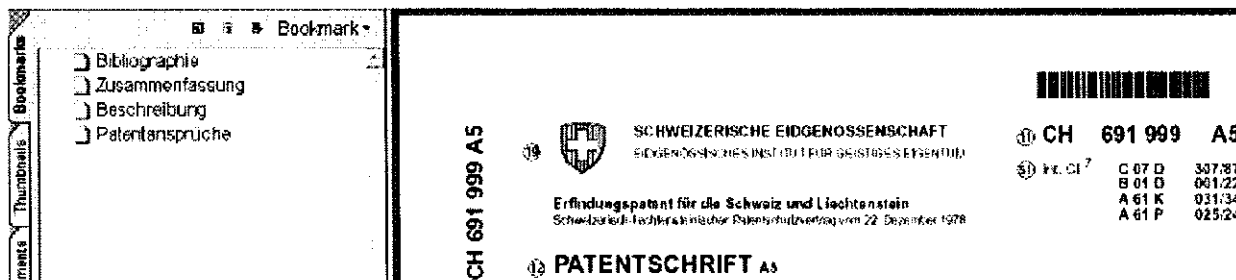
Erfindungspatent für die Schweiz und Liechtenstein
 Schweizerische Eidgenossenschaft, Patentschutzamt, 3003 Bern, Schweiz

CH 693 492 A5

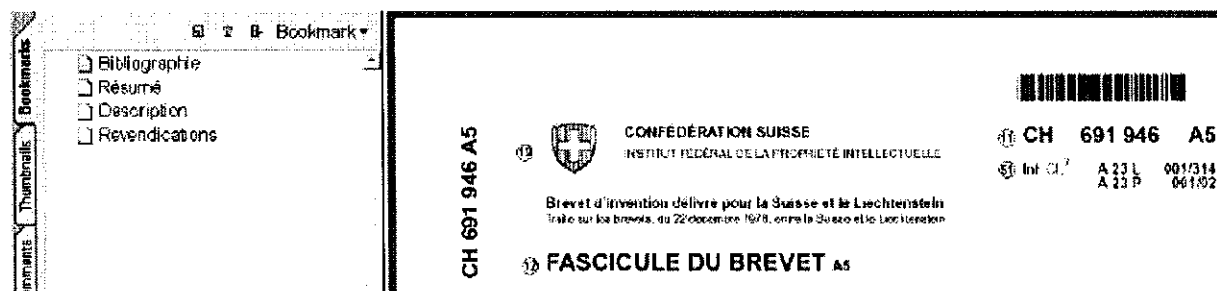
Int. Cl. 7 A 47 B 083/04

PATENTSCHRIFT A5

Esempio dalla Svizzera (in tedesco):



Esempio dalla Svizzera (in francese):



Consegna senza supporto fisico

Per la consegna delle immagini, l'EPO accetta File Transfer Protocol (FTP) o Web Services.

Per maggiori informazioni su una qualsiasi delle due opzioni sopra menzionate, non esitare a contattare il team di Front Office dell'EPO (frontoffice@epo.org), che sarà in grado di fornire ulteriori dettagli riguardanti il server FTP dell'EPO, la struttura delle cartelle FTP preferita dall'EPO e la convenzione di denominazione dei file e i requisiti minimi dell'EPO per i servizi web.

Correttezza, completezza, tempestività

Correttezza

indica l'accuratezza del formato e del contenuto dei dati:

- XML conforme agli standard e ben formato che esegue la convalida rispetto a un DTD o XSD distribuito in precedenza
- Formati numerici armonizzati e codici tipo ben documentati prima dello scambio per poter generare regole commerciali per la convalida automatica dei dati. Citazione, numeri di domanda e di priorità dovrebbero, ove possibile, essere forniti in formato

DOCDB e possono essere convalidati tramite l'OPS Rest Services (Number Service¹⁶)

- In alternativa, i numeri di domanda /priorità possono essere presentati in WIPO ST.13
- I numeri di brevetto utilizzati nelle citazioni devono essere presentati almeno nel formato: Codice Paese (CC, Country Code) seguito da un numero (numero variabile di cifre), se disponibile, seguito da un codice tipo (es. A, A3, B1)
- Le citazioni NPL dovrebbero essere presentate in forma strutturata quando possibile o in alternativa come una stringa separata da virgole
- Soglia massima del 5% per i difetti dei dati all'interno di una rispettiva consegna dei dati

Completezza

si riferisce alla copertura completa del frontfile/backfile e del 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 dovrebbe esserci idealmente una corrispondenza del 100% con l'**authority file**¹⁷ di tutte le pubblicazioni sin dall'inizio. Gli uffici nazionali sono responsabili della fornitura di un'authority file completa secondo la specifica
- Le citazioni e i dati full-text dovrebbero essere completi dal 1973 in poi

Informazioni pertinenti facoltative

- **File di contenuto per la consegna dei dati per ogni consegna**
Il file di contenuto per la consegna dei dati indica il numero totale di documenti inclusi in ogni nuova consegna e il corrispondente elenco di numeri di pubblicazione. Questo può essere visto come un aggiornamento regolare dell'authority file storica, se il formato è conforme alla specifica dell'authority file (vedi nota a piè di pagina 21)

Tempestività

viene rispettata quando ci sono consegne regolari e puntuali in giorni e tempi prestabiliti, in modo da rendere possibile il recupero e il caricamento automatico dei dati.

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

Qualità dati

L'Ufficio brevetti esercita la ragionevole diligenza per garantire l'accuratezza e la completezza dei suoi dati. Se si ritiene che tali dati debbano essere corretti, modificati o integrati, si comunicheranno tali correzioni, modifiche o integrazioni all'EPO senza indugio, se possibile, alla data di spedizione successiva.

Al ricevimento, tutti i nuovi dati saranno sottoposti a test approfonditi per la compatibilità e la conformità con i sistemi EPO. Una volta che l'EPO lo avrà approvato, verrà caricato

¹⁶Number Service OPS: <http://ops.epo.org/rest-services/number-service>

¹⁷ L'ultima specifica della file authority QaS, AF DTD e file XML di esempio AF possono essere acquisiti tramite Front Office EPO (frontoffice@epo.org)

secondo i criteri e le priorità stabiliti dallo stesso. Qualora l'EPO ritenga difettosi dei dati forniti dall'Ufficio brevetti, tali dati devono essere restituiti all'Ufficio brevetti per la verifica e, se necessario, la tempestiva correzione.

Eventuali modifiche alle specifiche e alle caratteristiche dei dati devono essere notificate all'EPO¹⁸ almeno sei mesi prima della loro attuazione. Tale notifica dovrebbe includere una specifica completa dei dati che includa un file di prova rappresentativo.

Sintesi

Per soddisfare i requisiti QaS, gli uffici IP dovrebbero mettere a disposizione dei loro partner di scambio, in modo tempestivo, completo e corretto, 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 immagine) in PDF con segnalibro
3. Dati citazione in WIPO ST.36
4. Dati full-text in WIPO ST.36
5. Trasferimento elettronico senza supporto fisico di tutti i dati sopra menzionati

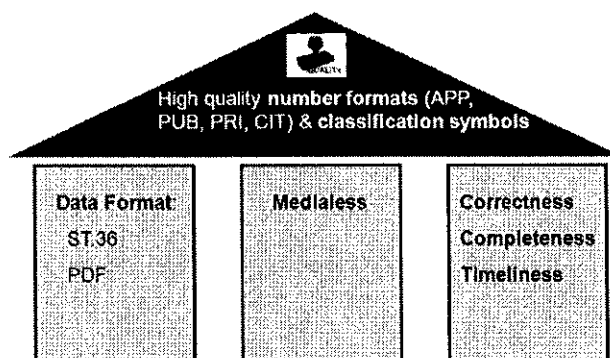
¹⁸ Le notifiche sulle modifiche devono essere indirizzate a frontoffice@epo.org

Allegato 1: Guida di riferimento rapido QaS

Quality at Source (QaS) incoraggia gli uffici IP ad assumere un ruolo attivo nel garantire una trasmissione di dati di alta qualità dei loro dati sui brevetti nazionali.

Si basa su quattro elementi fondamentali:

1. Dati bibliografici standardizzati:
 - a. Campi dati minimi
 - b. Numeri e date strutturati
2. Formato dei dati (WIPO ST.36 e PDF)
3. Scambio senza supporto fisico
4. Tempestivo, completo e corretto



1. Dati bibliografici standardizzati

- a. Campi dati minimi e numeri e date strutturati
- b. Formati numerici coerenti e codici tipo basati su normative WIPO per domanda/deposito, priorità, pubblicazione/registrazione, dati citazione
- c. Date standardizzate, corrispondenti al livello di pubblicazione corretto
- d. Informazioni strutturate sulla classificazione secondo le normative WIPO (ST8)

2. Formato dati

- a. Digitale, strutturato e conforme alle normative.
- b. **Normativa WIPO ST.36** per dati **bibliografici** (incl. riassunto, classificazione), (arricchite) **citazioni** e indicizzato **full-text** con immagini/formule incorporate.
- c. **PDF 1.3** o versione successiva, **segnalibro con indicizzazione** delle sottoparti del brevetto (ST.33) e delle radici di primo livello.

3. Scambio dati senza supporto fisico

- a. **Applicazione di trasferimento dati (servizi Web)** per l'invio automatico dei dati.
- b. **FTP** con la struttura concordata di directory/cartelle e convenzione di denominazione dei file.

4. Correttezza, completezza e tempestività

- a. **XML** conforme agli standard e ben formato in combinazione con specifico DTD.
- b. **Formati numerici e codici tipo ben armonizzati e documentati.**
- c. **Soglia massima del 5%** per eventuali difetti di dati all'interno di una rispettiva consegna.
- d. **Rettifica, modifica o integrazione dei dati** deve essere attuato e comunicato all'EPO (frontoffice@epo.org) **senza indugio.**
- e. **Corrispondenza al 100%** con l'**authority file** di tutte le pubblicazioni emesse per dati bibliografici e immagini (citazioni e full-text dal 1973 in poi).
- f. **Disponibilità di tutti i tipi di dati** (es. dati bibliografici, immagini, citazioni e full-text) per la stessa pubblicazione.
- g. Facoltativo: **File di contenuto per la consegna dei dati** per garantire la completezza di una consegna.

- h. **Consegne puntuali e regolari su giorni/tempistiche prestabilite il più vicino possibile alla data di pubblicazione effettiva.**
- i. **Tutte le modifiche devono essere segnalate e documentate almeno 6 mesi prima dell'attuazione.**

