

■ Quality Control in Terminology Management

Cristina Valentini
Head, Terminology Unit,
Support Section
PCT Translation Division

Geoffrey Westgate
Head
Support Section
PCT Translation Division

Copenhagen – June 22, 2016

World Intellectual Property Organization (WIPO)

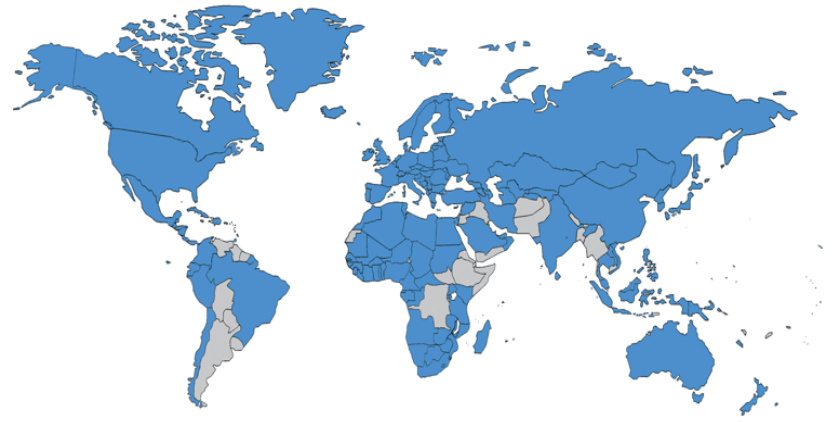
- Specialized UN agency for intellectual property (IP)

- Mission:
 - deliver global services for protecting IP
 - shape international IP rules
 - use IP for development
 - provide access to the world's IP information via
 - PATENTSCOPE
 - Global Brand Database
 - WIPO Lex
 - WIPO Pearl (since Sep 2014)



Patent Cooperation Treaty (PCT)

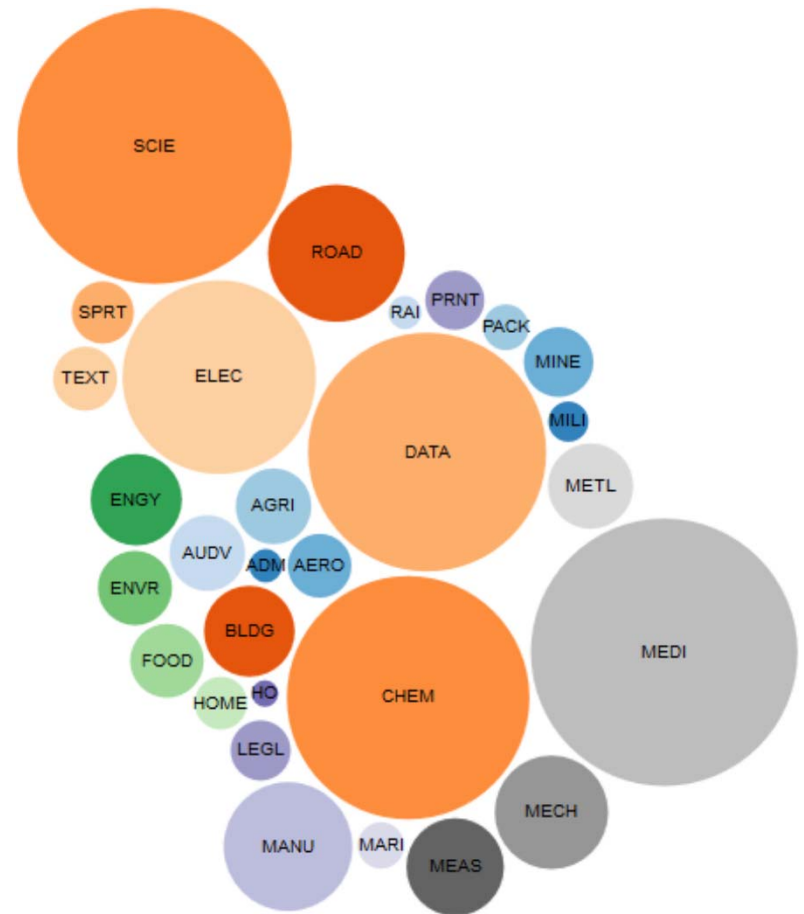
- treaty administered by WIPO
- unified procedure for filing patent applications to protect inventions in 149 contracting states
- 10 languages of publication
- improve access to patent applications via:
 - publication in PATENTSCOPE, and
 - translation of titles, abstracts and drawings into EN and FR, and patentability reports into EN (125+ million words of translation in 2015)



THE PCT TERMBASE

PCT Termbase: Contents

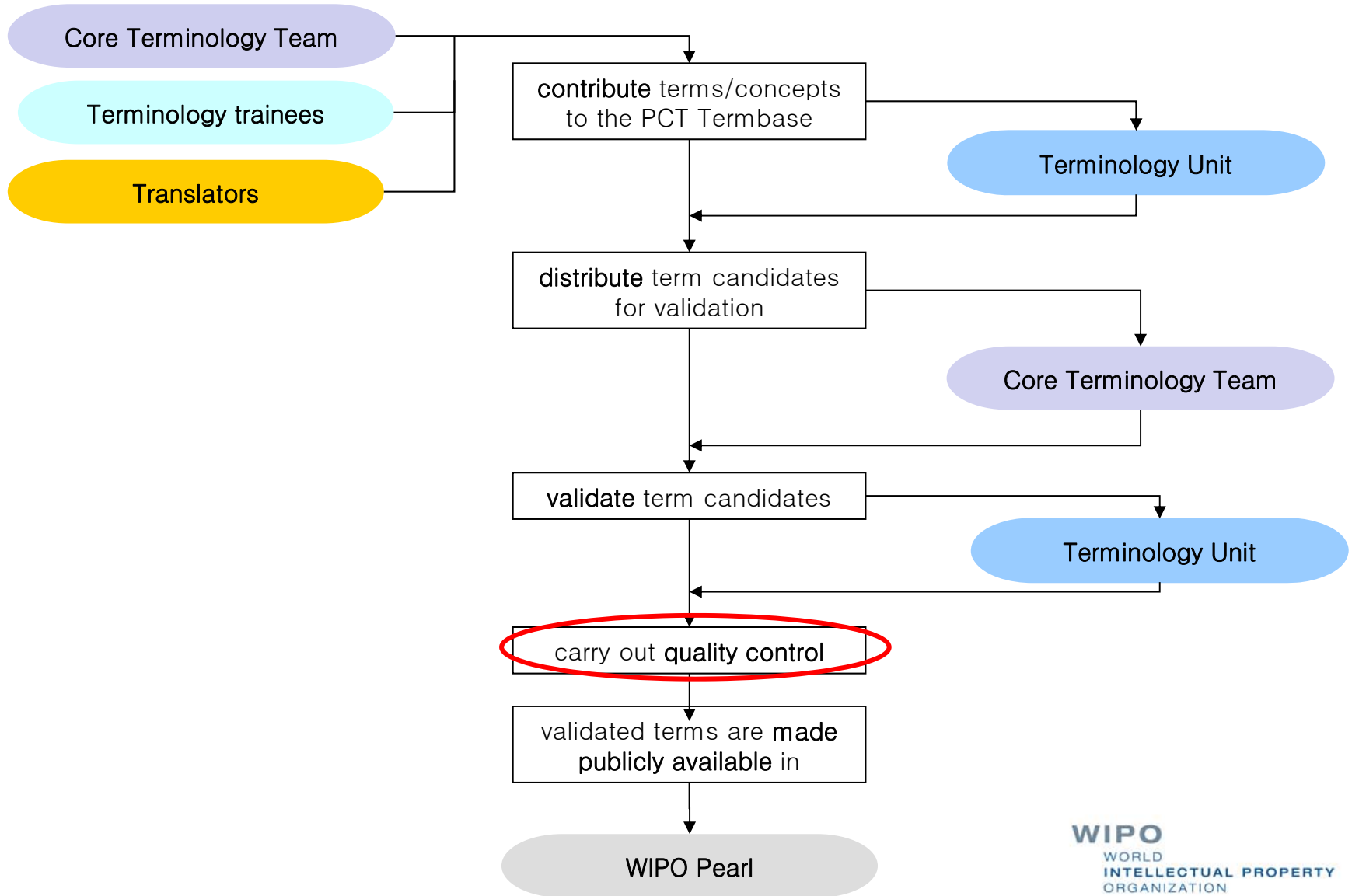
- Concepts and terms extracted from patents
- Ten PCT languages: AR, DE, EN, ES, FR, JA, KO, PT, RU, ZH
- 45 language combinations
- 29 subject fields (and 311 related subfields)
- 17,000 concepts
120,000 terms
12,000 concept relations



PCT Termbase: Main Features

- one concept, one record
- a record contains:
 - indication of subject field and subfield
 - concept relations (hierarchical/non-hierarchical)
 - designations in up to 10 languages
 - indication of preferred term vs. synonym
 - one (defining) context for each term
 - source (patent or academic publication)
 - reliability code for each term (1-4)
 - additional information, e.g. term usage
 - workflow status information (candidate, validated, unresolved)

Terminology Workflow



WIPO PEARL: THE PCT TERMBASE GOES LIVE

Watch the video at www.wipo.int/wipopearl

QUALITY CONTROL IN TERMINOLOGY MANAGEMENT

Why a QC procedure for terminology?

- To deliver a reliable terminology product to:
 - internal and external PCT translators to improve the quality/consistency of translations
 - the public at large so WIPO is seen as a reliable provider of multilingual scientific, technical and IP terminological information

Principles of Terminology QC

- A terminology database is assumed to be in compliance with the quality standards when it is «fit for purpose», i.e. suitable for the end use for which it is intended (EN 15038:2006)
- A terminology record is considered to meet the required quality standards when the information it contains is complete and reliable

Terminology QC: Criteria (1)

■ Relevance: suitable concepts and terms



Optoelectronics: <laser>, <wavelength>, <semiconductor>, and <SOLED> (stacked OLED), <AMOLED> (active-matrix OLED)



Appellations (Windows, JAVA), trademarks, lone adjectives, general descriptive terms (designating classes of objects): *e.g. means, device, apparatus, medium.*

Special language in patents

- WO/2007/112308 Multifocal **contact lenses**
- WO/2015/0866928 **Spectacles** associated with a hearing aid
- WO/2012/162239 Material for **eyewear** and eyewear structure
- US6848784 **Eyeglass device** having primary and auxiliary spectacle frames
- US20030088241 **Eyesight correcting apparatus**
- US20060103808 **Eyesight improving device**



Terminology QC: Criteria (2)

- Completeness: mandatory fields are completed
 - Subject field/subfield
 - Original Entry Language
 - Term
 - Term Status
 - Term Reliability (validators)
 - Usage Label
 - Term Type
 - Context
 - Source

Terminology QC: Criteria (3)

■ Correctness:

- «substantive» correctness; e.g. term equivalence; synonymy
- «formal» correctness: e.g. spelling/grammar, punctuation, compliance to citation standards for references

Terminology QC: Criteria (4)

■ informativeness:

- Sources are credible: e.g. written by native speakers, derived from patents, scientific and technical articles
- Context are defining or knowledge-rich (Meyer 2001):



The term **wireless network** generally refers to a telecommunications network whose interconnections between nodes are implemented without the use of wires. WO/2008/121974



Networked ISL lamp communicates with the central system ISH (2D) via a **wireless network** (2B) and Internet (2C) delivering all collected data related to atmospheric conditions at the microlocation as well as ISL status information. WO/2013/019135

Terminology QC: Methodology

- Manual QC of contributions (validation)
- Semi-automated QC of validated terms

Manual QC of contributions: Categories of Errors

Major errors vs. minor errors

«Acceptable» (A) vs. «Not acceptable» (NA) for publication in WIPO Pearl

- Term
- Context
- Source
- Mandatory fields
- Data integrity (Schmitz 2001)
- Proofreading

Manual QC of contributions: Example

Validation							
Week 21-2016							
Record Number	OEL	Language(s) to be	Validated on	QC rating	Error category		Error description
					Major	Minor	
5204	EN	Es only	21.03.2016	N/A	Context		Term appearing in Spanish context is "borna" instead of "borne". Moreover it is in the field of ROAD and not ELEC.
5420	EN	Es only	22.03.2016	A			
5562	DE	Es only	22.03.2016	A		Data integrity	Add "spelling variant" to synonym block.
6830	DE	Es only	21.03.2016	A		Data integrity	Preferred term should be "contacto de reposo".
6992	EN	Es only	21.03.2016	A		Data integrity	"abbreviated form" not selected; English full form not given in the Term Note.
7414	ZH	Es only	21.03.2016	N/A	Term		" <i>valvula electronica</i> " vs. " <i>válvula termoiónica</i> ": it seems as if there is a difference between thermoionic valves and phototubes, so I would not include the term as a synonym although it is the most common type i.e. a diode.
4313	DE	ES only	11.03.2016	A		Data integrity	Reference signs from contexts taken from patents removed.
4123	DE	ES only	11.03.2016	N/A	Source		Source is from a translated book.
4123	DE	ES only	11.03.2016	N/A	Context		Context is missing.
7015	EN	Es only	21.03.2016	A		Proofreading	Context contains typos.

Semi-automated QC of Terminology Validation

- Check formal completion of data fields
- Run weekly on 100% of validated data
- Incorrect association of data field values
- Errors are identified, corrected and tracked statistically
- Major errors affect the upload and display of data in WIPO Pearl

Semi-automated QC of Terminology Validation: Examples

Data category	Description
Term Reliability	Records in which the <i>Term Status</i> value is <i>validated</i> and the <i>Term Reliability</i> score is 2 or 4, or has not been completed
Usage Label	Records in which, within a term block, the <i>Usage Label</i> is <i>proposed term</i> and the <i>Context</i> field has been completed Records in which, within a term block, the <i>Usage Label</i> value is <i>proposed term</i> and the <i>Term Reliability</i> score is 3
Term Type	Records in which, within a language block, the <i>Term Type</i> value <i>head term</i> has not been selected or has been selected more than once
Term Description	Records in which, within a term block, the <i>Term Type</i> value is <i>head term</i> and the <i>Term Description</i> is <i>spelling variant</i>

Conclusions

- Increase reliability of terminological resources
- Identify training needs for contributors/validators

- But: resources for manual QC not always available (terminologists, native speakers, etc.)
- And: terminology management tools could help

WIPO Pearl: Perspectives

- Localization of GUI
- Seek feedback from world users
- Collaborations:
 - terminology students (FTI, Geneva; DCU, Ireland; ...)
 - technical experts in outside institutions
 - Termium
- Integrate collections from other areas of WIPO (copyright, brands, designs)
- Develop an online terminology training module for patent terminology
- Become the reference for multilingual patent terminology



WIPO Pearl: Contacts

wipopearl@wipo.int

Cristina Valentini, cristina.valentini@wipo.int

(Head, Terminology Unit, Support Section, PCT Translation Division)

Geoff Westgate, geoff.westgate@wipo.int

(Head, Support Section, PCT Translation Division)

James Phillips, james.phillips@wipo.int

(Director, PCT Translation Division)