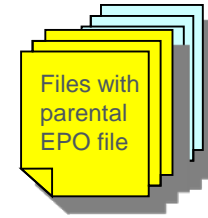# The EPO's experience in machine learning /automation
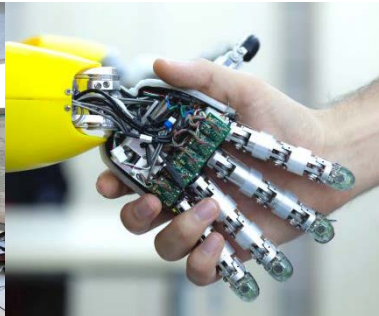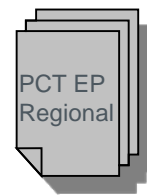
Machine Learning in pre-classification
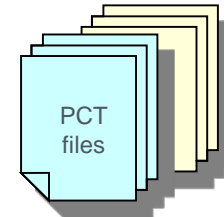# Routing files

Routing of files is done based on CPC or IPC (3 streams)

50% Classifications provided by other offices ie. CPC, IPC

PCT EP Regional

Nat. Search

Files with parental EPO file

2nd filings from GB, DK, AT, ES

50%

Intellectual preclassification 36%, quality 91%

EP files

PCT files
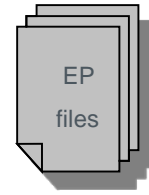
Automatic preclassification (Auto-Precla tool) 64%, quality 91%

EP files

PCT files
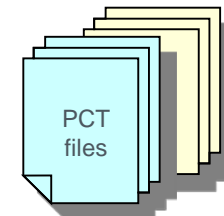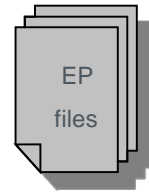
Auto-Preclassification

- To find the appropriate Technical Team: around 1400 CPC ranges

- Auto-Precla: analysing incoming applications and allocating a classification symbol for automated routing

- Taxonomy Engine provided by commercial company

Auto-Preclassification

EXTERNAL service

"Black Box"

Input, x

Output, F(x)

XML of
DESC, CLMS, ABS
+
title
+
reference to cited prior art

3 proposed classification symbols
with corresponding confidence levels

When to validate a CPC range

Training material: millions of documents classified in CPC

specific confidence level → minimum 80% stay in initial technical field

maximum 20% redistributed (Critical cases)
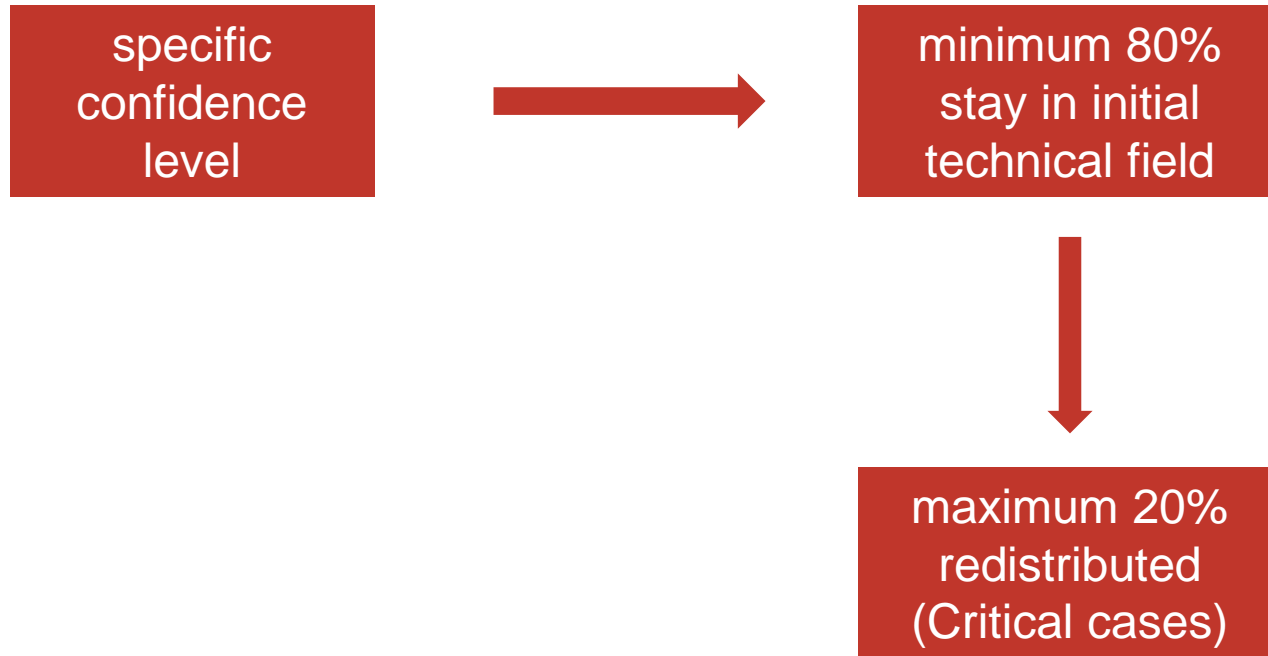
CPC Range "switches"

AutoPrec

External Service

EPO

"A61B17/00"

"0.67"

>

"0.65"

Direct

Directorates

Symbol

CPC Range

Confidence level [0 - 1]

EPO Threshold [0 - 1]

EPO has defined **for each CPC Range** a Threshold (i.e. minimum confidence value) to be achieved for each file

AutoPrec

External Service

EPO

"A61B17/00"

"0.59"

<

"0.65"

Directorates

Intellectual

Symbol
CPC Range

Confidence level [0 - 1]

EPO Threshold [0 - 1]

EPO has defined **for each CPC Range** a Threshold (i.e. minimum confidence value) to be achieved for each file

Automation in reclassification of documents in CPC (Auto-Recla service)

Load "Standard files" delivered by CPC revision process: Modified RCL, TRL, Quality

Modified "RCL"

(Revision Concordance List)

FROM	TO
A61M25/1018	A61M25/1018
A61M25/2014	A61M25/10181
	A61M25/10182
	A61M25/10183
	A61M25/10184
	A61M25/10185
	A61M25/10186
	A61M25/10187
■	■
A98G58/110	B63C57/032
A98G58/111	A98G58/110
A98G58/112	A98G58/111
A98G58/113	A98G58/112
A98G58/114	A98G58/113
A98G58/115	A98G58/114
A98G58/116	A98G58/115
A98G58/117	A98G58/11589
A98G58/118	A98G58/116
A98G58/119	A98G58/117
A98G58/120	A98G58/118
	A98G58/119
	A98G58/120

TRL

(Testing/Training Reclassification List)

B60R21/015	US2013004793
B60R21/01528	EP2548988
B60R21/01528	US2013022833
B60R21/017	US2013025366
B60R21/018	DE112012004434T
B60R21/019	DE202011104491U
B60R21/019	KR20140079472
B60R21/020	JPS5447162

QF

(Quality File)

To be achieved quality in %	**93,5**
Acceptable missed symbol(s)	0
Acceptable superfluous symbol(s)	1

Workflow

Classification
Quality Nominee

**Reclassify
remaining docs -
Send sample to
quality controller**

Auto-Recla Tool

**Reclassify e.g.
10% of docs more**

**Process complete
data and load into
database**

Sample
OK?

Yes

No

EPO classifier

DocDB

**Reclassify e.g.
20% of docs**

**RCL, TRL
Quality
files**

EPO classifier
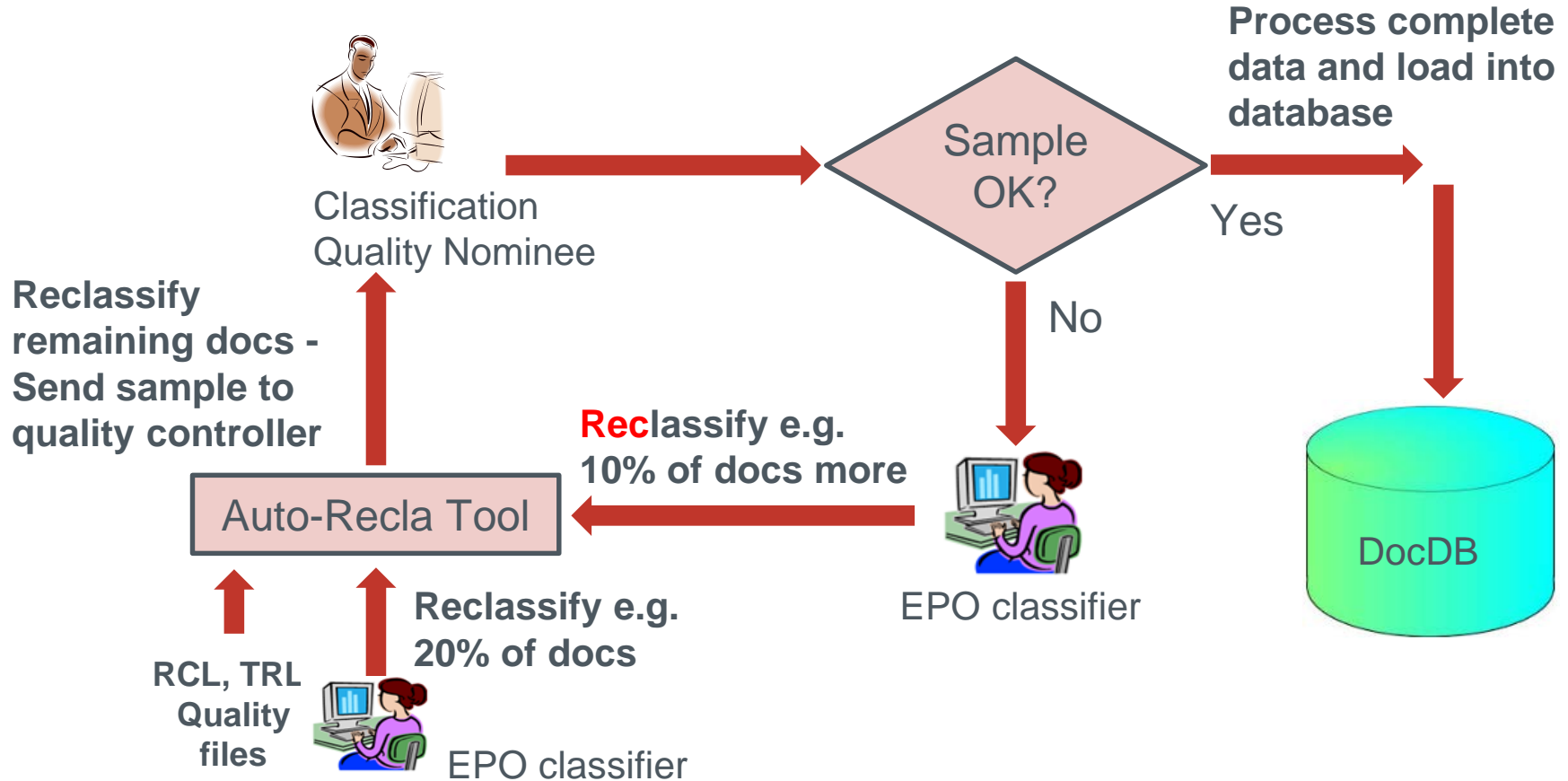
Quality of Auto-Recla

- Mandatory monitoring of quality!

- Achieved quality according to requirements set by Classification Quality Nominee

Machine Learning in Classification

Classification

- More difficult than pre-classification or reclassification

- All relevant symbols necessary for the search to be allocated (Invention and Additional symbols)

- Pre-class: 1400 ranges

- Classification: 260.000 symbols

- Choosing the right Artificial Intelligence software for classification

Machine Learning for Classification: 3 progressive levels

- Level 1: Artificial Intelligence only as support to the examiner/classifier

- Level 2: Artificial Intelligence as partial replacement to human classification

- Level 3: Artificial Intelligence as major replacement to human classification

Pilot of adding Machine Learning CPC symbols to documents classified in other patent offices
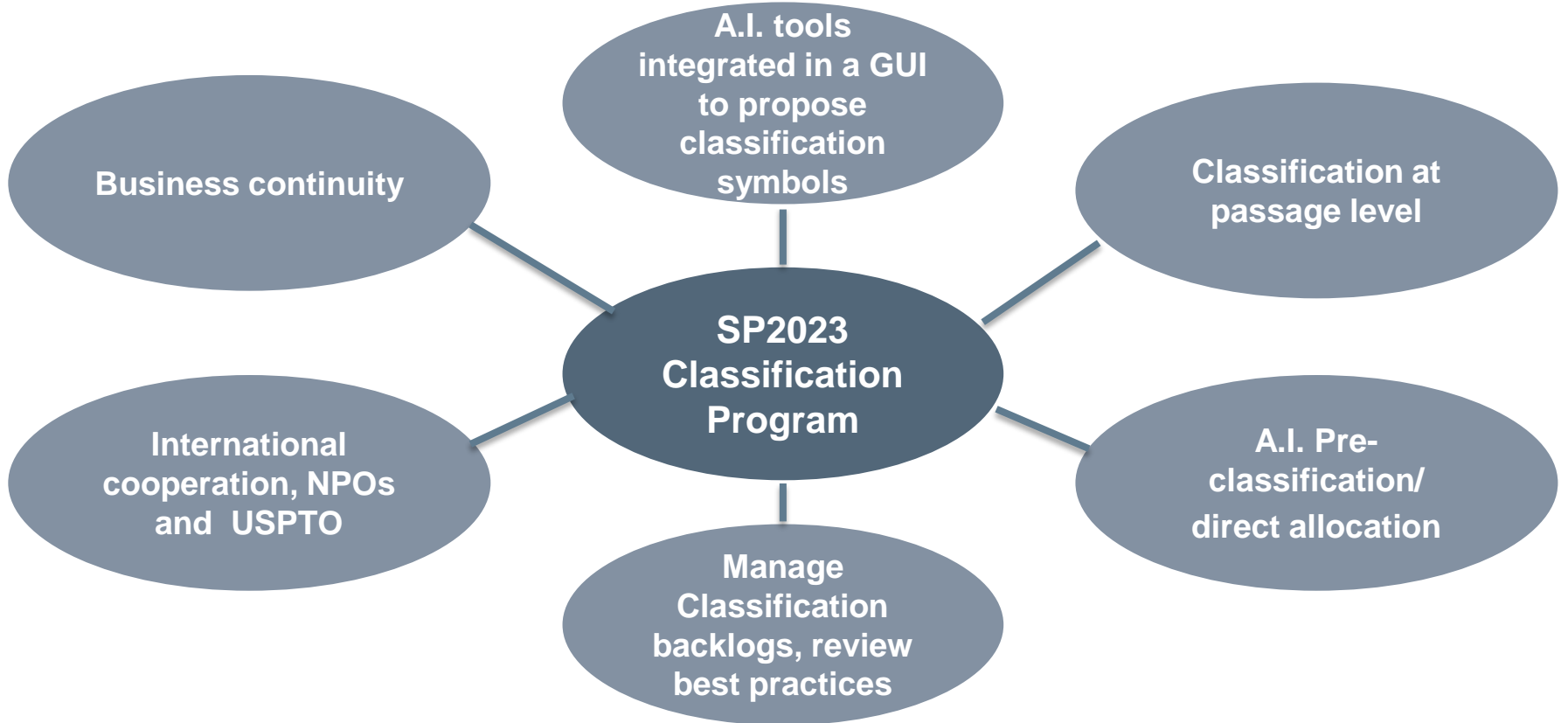
- In 21 technical fields
- Set of documents classified and previously checked by the Quality Nominee (QN)
- Machine Learning added 3 extra CPC symbols
- The set of documents with initial symbols + 3 extra symbols sent again to QN for quality check
- Comparison of results
- Outcome:
  - Increase of completeness
  - Some deterioration of correctness (noise)

Strategic Plan 2023

- **Master the Prior Art - Classification**

Strategic Plan 2023: Classification

A.I. tools integrated in a GUI to propose classification symbols

Business continuity

Classification at passage level

SP2023 Classification Program

International cooperation, NPOs and USPTO

A.I. Pre-classification/ direct allocation

Manage Classification backlogs, review best practices

Thank you!