Software for
Business Intelligence

BizInt Smart Charts

Patents & IP Sequences | Clinical Trials | Drug Pipelines

Challenges of Integrating Patent Information Resources

PIUG 2019 Annual Conference – Alexandria, Virginia

8 May 2019

John Willmore, BizInt Solutions

www.bizint.com

Outline

- Structural differences
- Patent families
- Content differences
- Special indexing

Structural Differences in Patent Resources

Method for reducing incidence of ascites in poultry by aid of **natamycin**

AB

(EP1143805)

Poultry such as chickens and turkeys are treated with **natamycin** to reduce overall mortality and to reduce mortality rates due to the disease ascites.

Protected countries

Granted: AU, MY, PT

Pending: MX

List of publications

Application number 1999US-09234703 Date 1999-01-21
US5985845 A - Patents Granted before 2001-04-15
PCT Application number 2000WO-US00188 Date 2000-01-06
WO200042864 A2 - International application published witho
WO200042864 A3 - Later publication of ISR with revised fron
Application number 2000CA-2360038 Date 2000-01-06
CA2360038 A1 - Application laid open
Application number 2000AU-0028462 Date 2000-01-06
AU2846200 A - Open to public inspection
AU773316 B2 - Patent proceeded by OPI
Application number 2000EP-0906870 Date 2000-01-06
EP1143805 A2 - Application published without search rep
EP1143805 A3 - Published search report
EP1143805 B1 - Patent specification
Application number 2000DE-6000914 Date 2000-01-06
DE60000914 D1 - Grant (no unexamined application
DE60000914 T2 - Trans. of EP patent
2003-07-10

US005985845A

United States Patent [19]

Carter

[11] **Patent Number: 5,985,845**

[45] **Date of Patent: Nov. 16, 1999**

[54] **METHODS FOR REDUCING MORTALITY RATES IN POULTRY**

[76] Inventor: **A. Franklin Carter**, 3001 Rockborough Ct., Ft. Collins, Colo. 80525

[21] Appl. No.: **09/234,703**

[22] Filed: **Jan. 21, 1999**

[51] **Int. Cl.**[6] **A61K 31/71**; C07H 17/08

[52] **U.S. Cl.** **514/31**; 536/6.5

[58] **Field of Search** 514/31; 536/6.5; 424/180

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,536,494	8/1985	Carter	514/31
4,600,706	7/1986	Carter	514/31

Primary Examiner—Gary L. Kunz
Attorney, Agent, or Firm—McDermott, Will & Emery

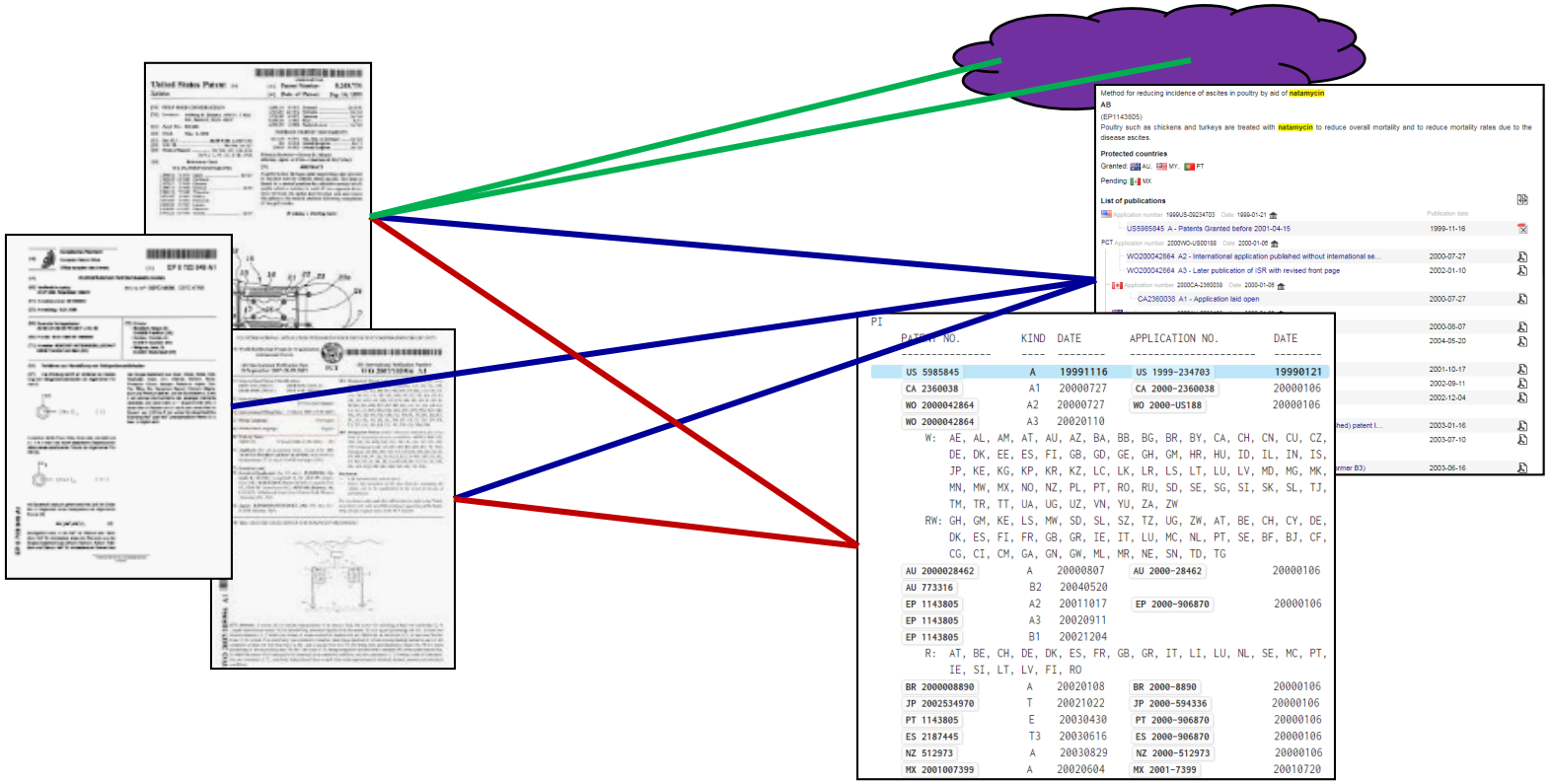
[57] **ABSTRACT**

Poultry such as chickens and turkeys are treated with natamycin to reduce overall mortality and to reduce mortality rates due to the disease ascites.

10 Claims, No Drawings

© 2019

Structural Differences in Patent Resources



© 2019 BizInt Solutions, Inc | www.bizint.com

Simple families in our portfolio

Bossa & Vikings

Olive

© 2019 BizInt Solutions, Inc | www.bizint.com

Extended family

Great Grandmother

Marta

Mother

Structural Differences in Patent Resources

- Structural differences are fundamental to the collections of patent resources
- Many of the systems are hybrids, organized both as family and publication collections

These differences appear in the user interfaces, in exports, and in the APIs of databases

© 2019 BizInt Solutions, Inc

Linking publications to families

- Method 1: linking based on family IDs
- Publications are assigned a family ID number by a publisher (e.g. Inpadoc or Derwent Family Number on Derwent Innovation)
- Records grouped by family ID by an aggregator (e.g. FSORT command on STN)

© 2019 BizInt Solutions, Inc

Linking publications to families

- Method 2: generating simple families
- “Build your own families”
- Need application or priority application number
- Normalize application numbers
- Sort

This technique is suitable for publication level data and simple families

Linking publications to families

- Method 3: linking publications to families
- Leverage the families constructed by publishers
- Normalize publication numbers
- Find the publication in a family and establish a link

© 2019 BizInt Solutions, Inc

Linking families to other families

- Use priority applications or publication numbers
- Normalize
- Build a transitive network linking records

Database	Common Family	Patent Family		
		Patent	Kind	Date
Derwent World Patents Index	US 2014356956	US20140356959	A1	20141204
Derwent World Patents Index	US 2014356956	US20140356956	A1	20141204
		[illegible]	A2	20141211
		WO2014197568	A3	20150312
		CA2914638	A1	20141211
FAMPAT	US 2014356956	~~US 2014356956~~	A1	2014-12-04
		~~US 2014356959~~	A1	2014-12-04
		~~US 9267135~~	B2	2016-02-23
GQPAT Gold+ Proteins	US 2014356956	US20140356959		20141204

© 2019 BizInt Solutions, Inc | www.bizint.com

Linking families to families - Strategies

- Link all members from all families
  - Transitive
  - Creates “super families” (union of all definitions)
- Link basic publications into a chosen family
  - Accuracy depends on controlling family being comprehensive
  - Unmatched records need to be handled
- Use extended family information
  (e.g. VLF IDs on PatBase)

© 2019 BizInt Solutions, Inc

Content differences: publication numbers

- Comparing publication numbers between systems can be challenging
- Number formats (years, leading zeroes)
- Embedded status or office information
- Implied authorities (“6,999,999”)
- Japan
- For some authorities, applications and grants use different serial number series CNxA = CNxB

© 2019 BizInt Solutions, Inc

Content differences: family presentation

- Patent Families are presented in fairly standard ways on most platforms
- Publication level collections often simply give a list of members

Database	Patent Family		
	Patent	Kind	Date
Derwent World Patents Index	US20140356959	A1	20141204
Derwent World Patents Index	US20140356956	A1	20141204
	WO2014197568	A2	20141211
	WO2014197568	A3	20150312
	CA2914638	A1	20141211
FAMPAT	US 2014356956	A1	2014-12-04
	US 2014356959	A1	2014-12-04
	US 9267135	B2	2016-02-23
GQPAT Gold+ Proteins	US20140356959		20141204
GQPAT Gold+ Proteins	US20140356956		20141204
PatBase	US 2014356959	A	2014-12-04
	US 2014356956	A	2014-12-04
	AU 2014274939	AA	2014-12-11
	WO 14197568	A2	2014-12-11
	WO 14197568	A3	2015-03-12
	CA 2914638	AA	2015-12-04
	KR 20160014036	A	2016-02-05

© 2019 BizInt Solutions, Inc | www.bizint.com

Content differences: family presentation

- Patent Families are presented in fairly standard ways on most platforms
- Publication level collections often simply give a list of members
- Sometimes family relationships are not exactly clear…
  Adis Drug Patents gives a list of US equivalents for non-US patents

WO 2007012793	
	US 9206135
	US 9643929
JP 5295765	
	US 9206135
	US 9643929
EP 2319836	
	US 9206135
	US 9643929

Why bother grouping by family?

- Comprehensive: Platforms with different search capabilities, indexing, timeliness
- Integrated: Linking unique content from different sources

Cortellis →

Orbit.com →

Enhanced Title	Indications	Patent Type	Classifications	Family Status			
				Pub No.	State	Status	Expiry
Monoclonal antibodies and vaccines against epitopes on the Ebola virus glycoprotein	Ebola virus infection	Product	Anti-Infectives Biologicals and Immunologicals				
				WO200116183	DEAD	LAPSED	2006-03-26
				AU7089600	DEAD	LAPSED	2006-03-26
				US6630144	ALIVE	GRANTED	2020-08-29
Monoclonal antibodies against glycoprotein of Ebola Sudan Boniface (ESB) virus - useful in the diagnosis and treatment of ESB virus infection.	Ebola virus infection	Diagnostic, Analysis and Assay Product (Macromolecule)	Anti-Infectives Biologicals and Immunologicals Diagnostics				
				WO2011071574	ALIVE	PENDING	2030-09-01
				EP2473525	DEAD	LAPSED	2014-08-27
				US2012164153	ALIVE	PENDING	2030-09-01
Ebola virus liposome vaccines	Ebola virus infection	Formulation	Anti-Infectives				

© 2019 BizInt Solutions, Inc

Why bother grouping by family?

- Comprehensive: Platforms with different search capabilities, indexing, timeliness
- Integrated: Linking unique content from different sources
- Analytics: Count each family once
- Surveillance: This publication is new, but have I already seen the family?

Content differences: Assignees

- Publishers must decide how to resolve differences between family members
- Varied levels of standardization
- Probable assignees, tracking reassignments
- Some content differences come from the source

At the Patent Office, we have the data we would need to normalize assignee names on the face, but these are legal documents so we can't.

Chris Harrison, IPO 2019 AVM, Nice, France

Content differences: challenges

- Mapping similar content into fields
  (e.g. titles – original, english, enhanced)
- Choosing among values from different records
  (e.g. preferring a particular vendor's inventors)
- Taking all values from different records
  (e.g. building composite patent families)
- Normalization

© 2019 BizInt Solutions, Inc

Content differences: Unique content

- Only rarely does a publisher have unique publications from the patent offices *[Citation required]*
- Unique structure (tagging)
- Value-added content (indexing, etc.)
- Application-specific content (e.g. gene sequences)

© 2019 BizInt Solutions, Inc

Content differences

CAS-9 - GenomeQuest, PatBase, DWPI (new STN), FAMPAT

	Title	Database	Patent Family: Patent	Kind	Date	Family Status: Pub No.	State	Status	Expiry	Probable Assignee	Sequence Locations: Seq. ID Number	% Identity	Length	Location	
1.	Modulating expression of a target nucleic acid comprises providing to the cell a guide RNA including a transcriptional activator or repressor domain as a fusion protein, and providing to the cell a nuclease null Cas9 protein	1.1 DWPI 1.2 DWPI 1.3 GPATPRT | link 1.4 GPATPRT | link 1.5 Patbase | link 1.6 FAMPAT | link	US 2014356959 US 2014356956 AU 2014274939 WO 14197568 WO 14197568 CA 2914638 KR 20160014036	A A AA A2 A3 AA A	2014-12-04 2014-12-04 2014-12-11 2014-12-11 2015-03-12 2015-12-04 2016-02-05	US 20140356956 A1 US 9267135 B2	ALIVE ALIVE	PENDING GRANTED	2034-06-04 2034-06-04	PRESIDENT AND FELLOWS OF HARVARD COLLEGE	US20140356959-0001 US20140356956-0001	100.00 100.00	1368 1368	probable disclosure (not found by automated parsing) probable disclosure (not found by automated parsing)	1.3 1.4
		1.1 DWPI			1.5 Patbase				1.6 FAMPAT	1.5 Patbase					
2.	New bacteriophage comprises polynucleotide expressing RNA-directed DNA-binding polypeptide comprising nuclease module, and targeting module comprising guide RNA, for restricting growth of host cell, and for preparing antiseptic composition	2.1 DWPI 2.2 DWPI 2.3 GPATPRT | link 2.4 GPATPRT | link 2.5 Patbase | link 2.6 FAMPAT | link	WO 15070193 US 2015132263 US 2015353901	A1 A A	2015-05-14 2015-05-14 2015-12-10	WO 201570193 A1 US 20150132263 A1 US 20150353901 A1	ALIVE ALIVE ALIVE	PENDING PENDING PENDING	2034-11-11 2034-11-11 2034-11-11	RADIANT GENOMICS INC	US20150132263-0002 US20150353901-0002	100.00 100.00	1368 1368	claim: 19; 20 claim: 19; 20	2.3 2.4
		2.1 DWPI			2.5 Patbase				2.6 FAMPAT	2.5 Patbase					

Common concepts…

	Title	Database	Patent Family		
			Patent	Kind	Date
1.	Modulating expression of a target nucleic acid comprises providing to the cell a guide RNA including a transcriptional activator or repressor domain as a fusion protein, and providing to the cell a nuclease null Cas9 protein	1.1 DWPI 1.2 DWPI 1.3 GPATPRT | *link* 1.4 GPATPRT | *link* 1.5 Patbase | *link* 1.6 FAMPAT | *link*	US 2014356959 US 2014356956 AU 2014274939 WO 14197568 WO 14197568 CA 2914638 KR 20160014036	A A AA A2 A3 AA A	2014-12-04 2014-12-04 2014-12-11 2014-12-11 2015-03-12 2015-12-04 2016-02-05
	1.1 DWPI				1.5 Patbase
2.	New bacteriophage comprises polynucleotide expressing RNA-directed DNA-binding polypeptide comprising nuclease module, and targeting module comprising guide RNA, for restricting growth of host cell, and for preparing antiseptic composition	2.1 DWPI 2.2 DWPI 2.3 GPATPRT | *link* 2.4 GPATPRT | *link* 2.5 Patbase | *link* 2.6 FAMPAT | *link*	WO 15070193 US 2015132263 US 2015353901	A1 A A	2015-05-14 2015-05-14 2015-12-10
	2.1 DWPI				2.5 Patbase

… unique content …

Family Status				Probable Assignee
Pub No.	State	Status	Expiry	
US 20140356956 A1	ALIVE	PENDING	2034-06-04	PRESIDENT AND FELLOWS OF HARVARD COLLEGE
US 9267135 B2	ALIVE	GRANTED	2034-06-04	
1.6 FAMPAT				1.5 Patbase
WO 201570193 A1	ALIVE	PENDING	2034-11-11	RADIANT GENOMICS INC
US 20150132263 A1	ALIVE	PENDING	2034-11-11	
US 20150353901 A1	ALIVE	PENDING	2034-11-11	
2.6 FAMPAT				2.5 Patbase

… and summarized content

Sequence Locations

Seq. ID Number	% Identity	Length	Location	
US20140356959-0001	100.00	1368	probable disclosure (not found by automated parsing)	1.3
US20140356956-0001	100.00	1368	probable disclosure (not found by automated parsing)	1.4
US20150132263-0002	100.00	1368	claim: 19; 20	2.3
US20150353901-0002	100.00	1368	claim: 19; 20	2.4

Special indexing

- Many types of indexing are structurally very simple – lists of terms or phrases assigned to a document.
- Controlled vocabulary for indices may come from a complex ontology, but the structure as applied to a document remains simple.

© 2019 BizInt Solutions, Inc

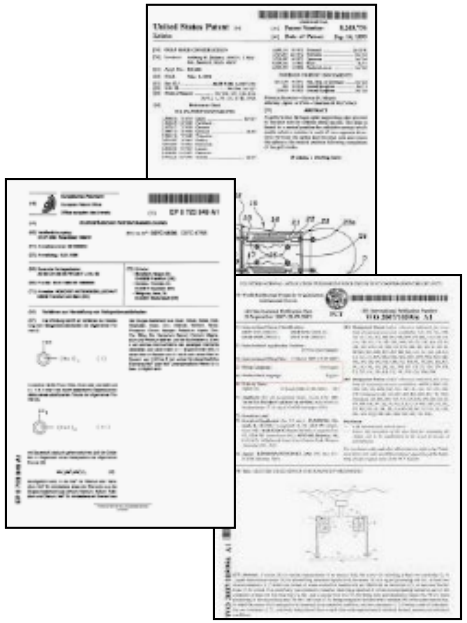
Special indexing - chemistry

- At 2018 PIUG Annual Conference we introduced an Index of Hit Structures from chemical structure searches in REGISTRY / CAplus on STN.

"Exemplified Compounds Table linked to Citing Publications“, Maddy Marley, 2018 PIUG

BizInt Workshop (bizint.com/slides)

Special indexing - chemistry



Method for reducing incidence of ascites in poultry by aid of **natamycin**

AB

(EP1143805)

Poultry such as chickens and turkeys are treated with **natamycin** to reduce overall mortality and to reduce mortality rates due to the disease ascites.

Protected countries

Granted: AU, MY, PT

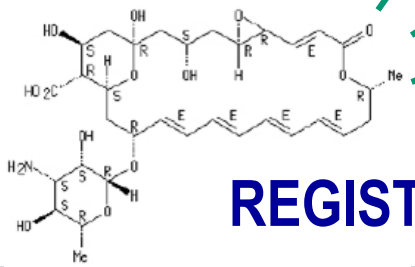
Pending: MX

List of publications

US5985845 A - Patents Granted before 2001-04-15		1999-11-16
WO2000042864 A2 - International application published without international se...		2000-07-27
WO2000042864 A3 - Later publication of ISR with revised front page		2002-01-10
CA2360038 A1 - Application laid open		2000-07-27
AU2846200 A - Open to public inspection		
AU773316 B2 - Patent proceeded by OPI		
EP1143805 A2 - Application published without search report		
EP1143805 A3 - Published search report		
EP1143805 B1 - Patent specification		
DE60000914 D1 - Grant (no unexamined application published) patent		
DE60000914 T2 - Trans. of EP patent		
ES2187445 T3 - Translation of granted European patent (former B3)		

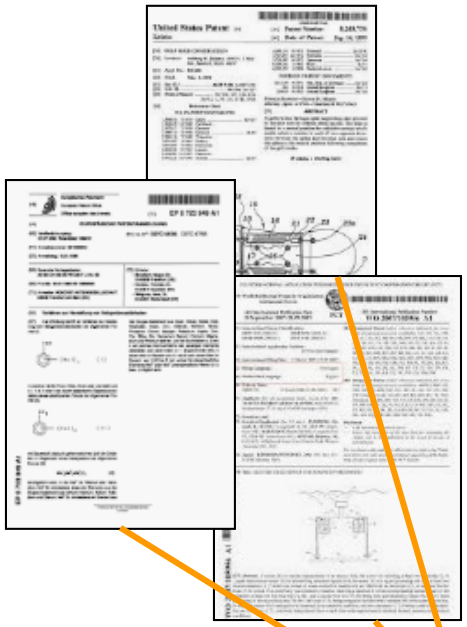
PI

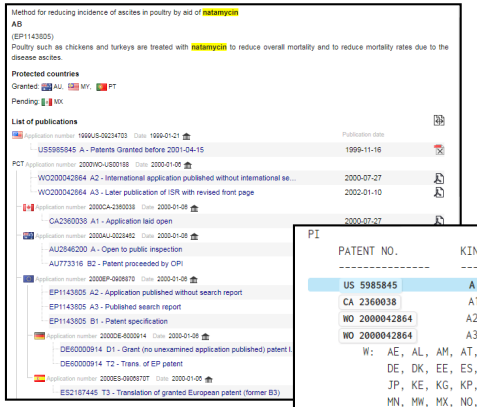
PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 5985845	A	19991116	US 1999-234703	19990121
CA 2360038	A1	20000727	CA 2000-2360038	20000106
WO 2000042864	A2	20000727	WO 2000-US188	20000106
WO 2000042864	A3	20020110		
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AU 2000028462	A	20000807	AU 2000-28462	20000106
AU 773316	B2	20040520		
EP 1143805	A2	20011017	EP 2000-906870	20000106
EP 1143805	A3	20020911		
EP 1143805	B1	20021204		
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				
BR 2000008890	A	20020108	BR 2000-8890	20000106
JP 2002534970	T	20021022	JP 2000-594336	20000106
PT 1143805	E	20030430	PT 2000-906870	20000106
ES 2187445	T3	20030616	ES 2000-906870	20000106
NZ 512973	A	20030829	NZ 2000-512973	20000106
MX 2001007399	A	20020604	MX 2001-7399	20010720



REGISTRY™

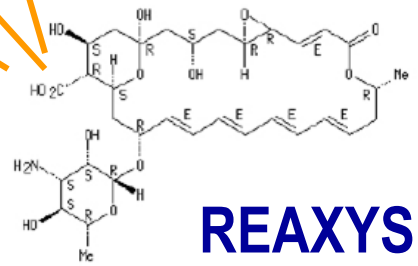
Special indexing - chemistry





PI

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 5985845	A	19991116	US 1999-234703	19990121
CA 2360038	A1	20000727	CA 2000-2360038	20000106
WO 2000042864	A2	20000727	WO 2000-US188	20000106
WO 2000042864	A3	20020110		
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AU 2000028462	A	20000807	AU 2000-28462	20000106
AU 773316	B2	20040520		
EP 1143805	A2	20011017	EP 2000-906870	20000106
EP 1143805	A3	20020911		
EP 1143805	B1	20021204		
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				
BR 2000008890	A	20020108	BR 2000-8890	20000106
JP 2002534970	T	20021022	JP 2000-594336	20000106
PT 1143805	E	20030430	PT 2000-906870	20000106
ES 2187445	T3	20030616	ES 2000-906870	20000106
NZ 512973	A	20030829	NZ 2000-512973	20000106
MX 2001007399	A	20020604	MX 2001-7399	20010720



REAXYS

REGISTRY™

THE JOURNEY CONTINUES...

doxifun.com/puppies

Software for Business Intelligence

BizInt Smart Charts

Patents & IP Sequences | Clinical Trials | Drug Pipelines

Questions?

john.willmore@bizint.com