

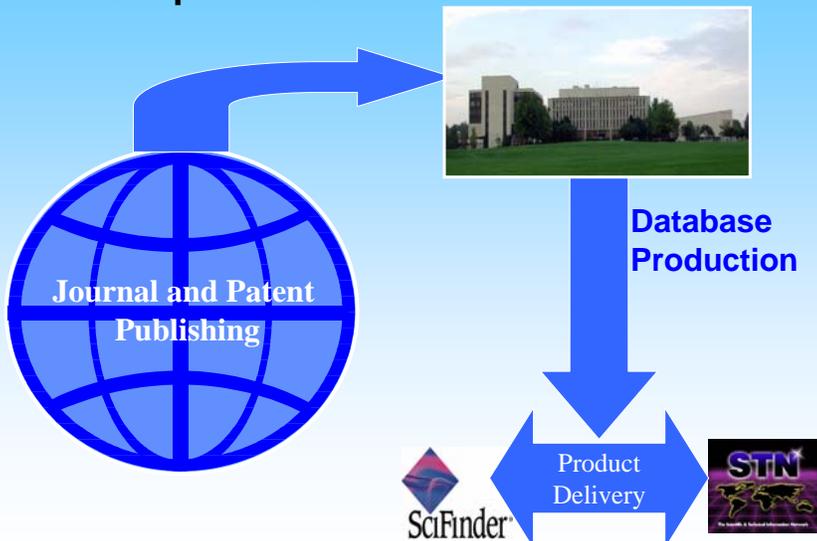
# Harvesting the scientific information...: what non-patent specialists should know.

230<sup>th</sup> American Chemical Society  
National Meeting, August 28, 2005

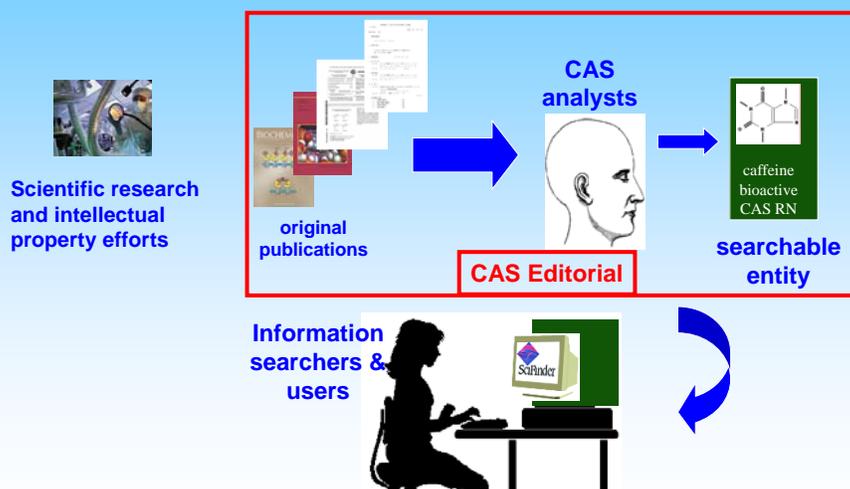
William M. Mercier, Jan Williams, CAS



CAS (a division of the ACS) creates databases from published worldwide information



## CAS scientists analyze and indexing scientific literature



## Data from CAS can be used to:

- Assist in making business critical decisions
- Direct a research project
- Assess the prior art for patentability



## Search output should be evaluated with the following criteria in mind

- Content of the database(s)
- Quality of the data therein
- Currency/timeliness



## Business insight question

- *Did DuPont's patenting habits shift after a pronouncement by CEO Chad Holliday in 1999 (re)affirming DuPont's commitment to biotechnology?\* Is DuPont patenting more in biotechnology?*

*\*Chemical & Engineering News  
September 27, 1999*



## A scenario/approach

1. Access the CAPlus<sup>SM</sup> database on STN<sup>®</sup>
2. Isolate DuPont patents published after 1990
3. Tally the output by year and by subject matter
4. Compare number of patents in biotechnology vs. other areas over time



## Results can be easily interpreted via STN's TABULATE command

IC	Publication Year														
	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991
C08	454	790	761	756	663	591	614	526	471	468	474	424	353	334	240
C07	237	419	485	509	540	493	457	388	382	365	344	336	326	279	185
C09	155	231	211	196	173	185	189	189	158	158	160	131	129	82	66
B32	135	233	187	112	118	101	110	79	60	70	67	54	58	56	41
G03	65	98	106	98	96	96	95	77	81	99	103	102	104	95	102
H01	147	171	134	97	77	86	67	67	65	61	50	55	46	54	42
<b>C12</b>	<b>130</b>	<b>171</b>	<b>177</b>	<b>169</b>	<b>200</b>	<b>183</b>	<b>120</b>	<b>75</b>	<b>54</b>	<b>58</b>	<b>60</b>	<b>55</b>	<b>59</b>	<b>45</b>	<b>32</b>
A01	68	82	98	75	82	91	112	110	98	87	107	90	96	83	50
A61	50	98	131	149	161	139	116	94	80	101	98	98	91	78	64
D01	50	103												62	57

IC = International Patent Classification Code (IPC)

C08 is polymers

C07 is organic chemistry

C09 is dyes/paints

**C12 is biochemistry**



## Conclusions

- DuPont is patenting more in this decade in biotechnology than in the last decade
- With regard to Dupont's traditional technologies, the number of patent documents has doubled; but the number of patent documents in biotechnology has tripled



## Confidence question: What confidence do I have that all relevant documents are retrieved?

- **What you should know**
  - CAS covers all documents in chemistry, biochemistry and chemical engineering
    - Patents from an authoritative list of IPCs are included: <http://www.cas.org/EO/caspat.html>
  - Patents from 50 active issuing authorities are included and indexed
  - All patents include the IPC indexing



## Confidence question: What confidence do I have that all DuPont patents are retrieved?

- **What you should know**
  - The CAS Company Name Thesaurus ensures that all variations of the company name (there are over 2100 variations of DuPont!) are found
  - The Thesaurus helps the user control the inclusion or exclusion of joint ventures (e.g. DuPont-Dow Elastomers), former corporate entities (e.g. DuPont Pharma) or add geographical precision (e.g. DuPont Japan)



## Research objective

- *Lately there is daily news regarding the fuel additive methyl tert-butyl ether (MTBE). Research continues to find a replacement for MTBE. What substances show promise as candidate replacements?*



## A scenario/approach

1. Access the CAPLUS and CAS REGISTRY<sup>SM</sup> databases on STN
2. Isolate literature pertaining to fuel additives for increasing octane number or improving mileage
3. Extract chemical substances mentioned specifically as fuel additives
4. Limit to substances that are similar in some physicochemical way(s) to MTBE: carbon count, molecular weight, flash point, boiling point, heat of vaporization, vapor pressure, density, etc.
5. Prepare a table of these substances



## Report-type/tabular output can be generated using STN EXPRESS

Possible MTBE replacements

CAS Number®	Name	Molecular Formula	Graphics	Reference(s)
62108-41-2	Pentane, 2-methoxy-2,4,4-trimethyl	C <sub>9</sub> H <sub>20</sub> O		IN FILE CA 29 IN FILE CAPLUS 29
17860-39-8	Propanoic acid, 2-methoxy-2-methyl-, methyl ester	C <sub>6</sub> H <sub>12</sub> O <sub>3</sub>		IN FILE CA 40 IN FILE CAPLUS 40 IN FILE CAOLD 3
6795-88-6	Pentane, 2-methoxy-	C <sub>6</sub> H <sub>14</sub> O		IN FILE CA 33 IN FILE CAPLUS 33 IN FILE CAOLD 4
4098-71-9	Cyclohexane, 5-isocyanato-1-(isocyanatomethyl)-1,3,3-trimethyl	C <sub>12</sub> H <sub>18</sub> N <sub>2</sub> O <sub>2</sub>		IN FILE CA 4725 IN FILE CAPLUS 4734 IN FILE CAOLD 2
3852-11-7	Propanoic acid, 3-methoxy-2-methyl-, methyl ester	C <sub>6</sub> H <sub>12</sub> O <sub>3</sub>		IN FILE CA 82 IN FILE CAPLUS 82 IN FILE CAOLD 5
2110-78-3	Propanoic acid, 2-hydroxy-	C <sub>5</sub> H <sub>10</sub> O <sub>3</sub>		IN FILE CA 262 IN FILE CAPLUS 262

## Conclusions

- There are almost 100 substances that are candidates for further investigation
- These substances have been mentioned in the journal and patent literature as fuel additives for increasing the octane number or improving mileage
- These substances have physical and chemical properties specifically controlled by the search query



## Confidence question: What confidence do I have that all relevant substances are retrieved?

- **What you should know**
  - CAS covers all new chemistry of substances mentioned in journals, patents, technical reports, dissertations and other document types
  - CAS indexes substances in examples or specifically claimed in patent documents
  - CAS assigns ROLE indicators to describe the context of the chemical substance (e. g. MODIFIER OR ADDITIVE)
  - CAS assigns modifying terms to describe specific concepts mentioned in the document (e.g. octane or mileage)



## Confidence question: What confidence do I have that all relevant candidate substances are isolated?

- **What you should know**

- CAS's REGISTRY file has ~.5 billion property data points
- Property data in the CAS databases comprises:
  - Experimental properties: data points as indexed by CAS analysts and other sources
  - Calculated (predicted) properties
  - Tagged properties: documents indexed by CAS analysts point to references with data
- CAS's search services (STN and SciFinder®) offer extensive numeric search capabilities



## Surveying the prior art for assessing patentability

- *There's lots of news lately about athletes and performance-enhancing drugs. A researcher who is looking to develop and patent a new analytical technique for these substances would need to check for novelty as compared to all existing inventions and prior art.*



# A scenario/approach

Note: This scenario will focus on stanozolol and stanozolol-like substances

1. Access the REGISTRY database on STN or SciFinder. Run a substructure search to isolate all stanozolol derivatives
2. Isolate literature pertaining to analytical studies of these substances.
3. Examine via a visualization-type output a snapshot of this body of literature



## Results can be visualized via SciFinder's Panorama feature

Microsoft Excel - Book2

File Edit View Insert Format Tools Data Window Help

Type a question for help

B2 42

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1		Urine analysis	Mass spectrometry	Gas chromatography	Forensic analysis	Tandem mass spectrometry	Liquid chromatography	HPLC	Blood analysis	Pharmaceutical analysis	Extraction	Food analysis	TLC (thin layer chromatography)	Electrospray ionization mass spectrometry	Reversed phase HPLC	Calibration	Chemical ionization mass spectrometry	UV and visible spectroscopy	TLC (thin layer chromatography), high-performance	Chromatography	Sample preparation	Standard substances, analytical	Spectrophotometry	Radiochemical analysis	Solvent extraction	Bioassay
2	Stanozolol	42	30	26	22	11	9	11	9	8	7	5	4	2	3	1	2	2	2	2	1		2	2	2	2
3	3-Hydroxystanozolol	26	13	13	17	6	3	4	1		2			1		1				1	1	1	1		1	
4	Hormones, animal	23	19	18	23	11	11	1	1		7	3	1	4	1	1	1			2	1	2	1		1	
5	Methyltestosterone	20	14	16	18	8	5	6	4	3	4	3	1	2	2	1	1	2	2	2	2	1	2	2	2	1
6	Testosterone	20	19	18	19	5	4	5	3	2	3	3	1	1	2	2	1	2	2	2	1	1	2	1	1	2
7	Nandrolone	19	20	18	12	6	4	5	4	4	5	4	1	2	2	1	1	1	2		1	1		2	1	1
8	Boldenone	20	19	20	15	8	5	4	2	3	3	3	1	2	2	2	1	1	2	1	1				1	2
9	Fluoxymesterone	20	12	12	13	7	7	3	3	1	3	1	1	3	1	2	1	1	1	1	1	1		1		1
10	Methenolone	17	11	13	12	4	1	1						1	1	1				1	1		1			
11	Metandienone	16	12	12	12	6	5	4	3	3	1	2	1	2	1			2	1	1	1			1	1	

## Conclusions

- There is an abundance of literature (patents and journal articles) on analytical techniques for stanozolol and stanozolol-like substances
- There are opportunities to develop analytical techniques in this area



## Confidence question: What confidence do I have that all relevant references are retrieved?

- **What you should know**
  - CAS covers virtually all chemical and chemically-related documents from 1900 to the present
  - CAS monitors virtually all chemical substances. The REGISTRY file currently contains over 26 million organic substances and over 56 million sequences



## Confidence question: What confidence do I have that all relevant substances are retrieved?

- **What you should know**

- CAS search tools (STN, SciFinder, etc.) allow powerful substructure searching to find analogs [about 100 stanozolol derivatives were isolated]
- CAS uses controlled vocabulary (described in the CA Lexicon) to index concepts found in scientific documents and links these concepts to the substances being studied.



## Confidence question: What confidence do I have that my search results are up-to-date?

- **What you should know**

- The CAPLUS database provides bibliographic entries from about 1800 core chemistry/biochemistry **journals** within 2 days of receipt. An additional 7000 journals are processed as quickly as possible
- The CAPLUS database guarantees the bibliographic entries for **patent documents** from 8 major patent authorities within two days. Complete indexing for these documents is available within 27 days. An additional 42 authorities are processed as quickly as possible



**Because of CAS's dedication to coverage, quality, and currency, data from CAS can confidently be relied upon to:**

- Assist in making business critical decisions
- Direct a research project
- Assess the prior art for patentability



**Harvesting the scientific information...: what non-patent specialists should know.**

**230<sup>th</sup> American Chemical Society  
National Meeting, August 28, 2005**

**William M. Mercier, Jan Williams, CAS**

