My memories of Markush DARC: Anglo-French work at the leading edge

By Philippe Borne, INPI, 2020

I graduated from the University of Pharmacy in January 1985, without any precise idea where to orient my professional life. My somewhat anti-conformist image did not make me a good candidate for the industry. Selling drugs in a pharmacy seemed quite a boring activity. I tried a hospital lab, but analyzing blood samples left no memorable impressions either. In February 1986 the guidance counsellor of the university gave me a ring: "hey, they are on the lookout for two staff at INPI, you should apply".

"INPI, what is INPI" I thought. Intellectual Property was part of the law course, about 4 hours were dedicated to patents, trademarks … and other boring stuff for a young guy fond of chemistry, biology and other scientific matters: neither had IP left me a good impression. But I had no other routes to finding a job; I therefore went to the interview, and finally, thanks to my understanding of German, was hired. That's how I entered the Markush Wonderland.

Dr. Eugene A. Markush (1888-1968) Founder of Pharma Chemical Corporation - Photo: 1947

The man behind the name

Dr. Eugene A. Markush was born in Budapest in 1888, attended various Hungarian universities, and obtained a Ph. D. degree. He emigrated to America in 1913 and did post-graduate work at Columbia University in New York. He was employed as a chemist by Lederle Laboratories.

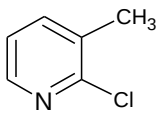
In 1917 Dr. Markush founded the Pharma Chemical Corporation of Bayonne, New Jersey and began the manufacture of pharmaceuticals, chemical intermediates and dyes.

In 1923, he filed a patent application for a method of preparing pyrazoline dyes usable for wool and silk, in which he claimed generic chemical structures in addition to those actually synthesized. Claim 1 of this application read:

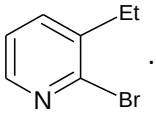
The process for the manufacture of dyes which comprises coupling with a halogen-substituted pyrazolone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of aniline and halogen substitution products of aniline.

This claim was challenged as being too unspecific. On appeal, the U.S. Commissioner of Patents ruled on the propriety of such claims. The patent was granted in 1924 as US 1,506,316. As it was the first ever granted patent claiming such generically defined structures, this type of structure became known as "Markush structures". The "Markush Doctrine" of patent law greatly increases flexibility in the preparation of claims for the definition of an invention.

What is a generic chemical – or a Markush – structure?

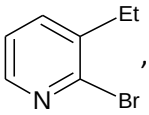
In real life, R&D departments invent this style of compound (compound A): 

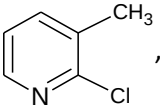
(this is a theoretical, and simple, example, just to make your understanding easier, especially if you are not a chemist). All the components of this compound are defined: we know that we have a 6-member ring having 3 single bonds and 3 double-bonds, one N atom (the other, non-apparent, atoms are carbon atoms). This ring has 2 substituents: a Cl – which stands for chlorine – atom, and a CH3 (a carbon having 3 hydrogen atoms) group.

Nobody would use this type of representation in a patent claim. Simply because, after reading such a claim, a competitor could for example file a patent application claiming this other compound (compound B) : 

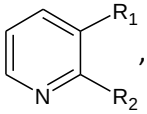
. It looks different, but it is actually very close to the previous one: Br – bromine – and Cl – chlorine – are both halogens (together with iodine and fluorine). The Et group – Et stands for ethyl – is very close to CH3, it just has one more carbon atom, the drawn-out form of Et being –CH2-CH3. Compound B is new with respect to compound A.

At that stage it should be remembered that in order to be patentable an invention must be novel, and involve an inventive step. Novelty alone is not enough. "Involve an inventive step", what does it mean really? American people are clearer, instead of saying that the invention "must involve an inventive step", they say that it should not be obvious. In Europe we say that "An invention shall be considered as involving an inventive step if, having regard to the state of the art, it is not obvious to a person skilled in the art" (European Patent Convention, article 56).

As a consequence of all this, in order to get a patent for 

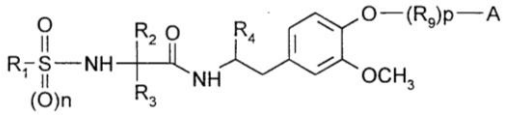
, the applicant needs to demonstrate that, "having regard to the state of the art", namely having regard to 

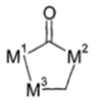
, it is not obvious to a person skilled in the art that the fact of replacing CH3 by Et and Cl by Br will enable an improved outcome. If the applicant can demonstrate that, the two main conditions for patentability (novelty and inventive step) would be satisfied. And, unfortunately for the inventor of the first compound, a patent would be granted covering the second one.

In order to avoid this pretty embarrassing outcome, most patent applicants in the organic chemistry area tend to file patents claiming the single compound of interest, and the family of said compound as well. They file this: 

, where R1 and R2 are not real atoms, they are groups representing a series of listed alternatives. That is what we call a Markush compound. Say that in this example, R1 = -CH3 and Et and even all the carbon chains; and R2 = -Cl; -Br, and all the halogens as well.

Patent offices accept this style of claim provided it is supported by a sufficient number of examples, that is to say a sufficient number of specific compounds, covered by the claim, described together with their process of preparation and all their physico-chemical characteristics. In other words, applicants must show that they really prepared the claimed compounds, and that their claim is not only a theoretical representation, never turned into practice.

In real life, claimed formula look like this : 

or even this 

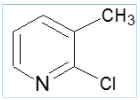
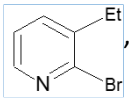
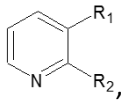
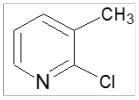
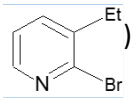
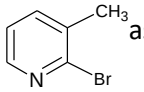
(out of 6 atoms, 3 are variables, M1, M2 and M3), where the level of variability can be very high.

In summary, a Markush structure is a shorthand used for describing sets of specific molecules with common features. It is a "structure with R-groups", also known as generic chemical structures. Their use in patents to protect whole classes of compounds with common properties was originally authorized by the ruling in 1925 by the US Patent Office.

Since then it has been allowed by all the patent offices in the world.

As a result of all this, a patent in the chemistry area usually has 3 styles of compound:

- specifically described examples,
- Markush claims …. and another odd species:
- so-called prophetics (*see diagram next page*).

If, starting from a list of exemplified compounds, say  and  , you elaborate a claim looking like  , (where R1 = -CH3; -ET; -carbon chains; and R2 = -Cl; -Br; -Halogens) just to gain a better protection of your invention, you can thereafter go the reverse way, and starting from your claim try to indicate all the compounds it covers. This reverse way is called enumeration. Of course you get back the 2 original examples ( and ); but you get  as well, not part of your examples, but embedded in your claim. That is what we call a “prophetic”. In the patent law, this type of compound is of great importance: they cannot take away the novelty of a future, identical, compound object of a latter patent; but they can seriously hamper the freedom to use such a future compound. Having the ability to search for those prophetics embedded in Markush claims is therefore key as part of a freedom to operate (FTO) search.

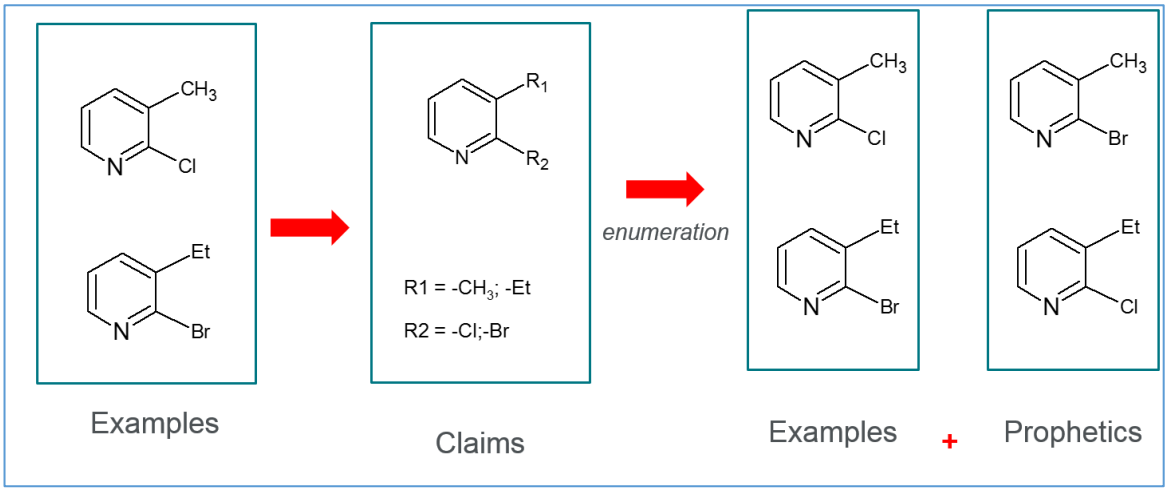


Figure 1 : specifics and prophetics

The fragmentation approach

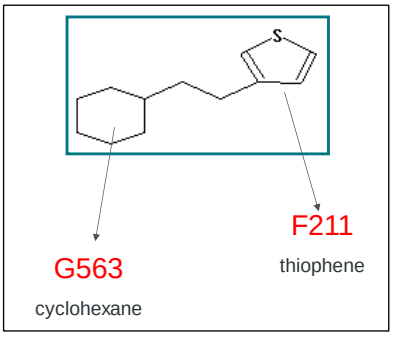
It can easily be understood that in the early 60s, running a prior art search in the chemistry area was a little bit like looking for a needle in a hay stack.

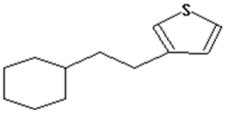
Not only had one to search in the millions of specific compounds synthesized by all the chemical and pharmaceutical companies on earth, but in the thousands of Markush formula claimed in the world as well, each of which can cover millions of specific compounds.

The only way to cope with that tricky challenge was to wade through the shelves of patent libraries.

Fortunately, Monty Hyams and Derwent had launched in 1963 the fragmentation coding system which together with the punch cards permitted to alleviate the pain by semi-automating the search process.

The principle behind the fragmentation codes is fairly simple: a compound is split into fragments, a code taken from a thesaurus is assigned to each fragment, and the codes related to the fragments diverted from the compound are stored in the punch card produced for said compound. In the adjacent picture, the G563 code represents the cyclohexane ring, and the F211 code represents the thiophene ring.



At searching, in order to pick up all the records relating to the 

compound, one simply need to search for all the records in the database (for all the punch cards, in the punch card era) having the G563 and F211 codes. Fragmentation codes were made searchable online in the World Patent Index on the Orbit host in 1976.

With no contest, this was a great progress as compared to the hiking tour in the patent library. In 1963 it is not exaggerated to say that it was revolutionary.

Some years later however, it turned out that the fragmentation code system had some drawbacks resulting from the fact that the topology was not stored. When searching for those molecules having both fragments A and B, you can pick up A-B-C, and A-C-B as well. Valid hits were obtained together with many false-drops, and the system proved too inaccurate. By the end of the 70s, there was a clear need for a better system. This did not prevent Derwent from carrying on indexing the fragmentation codes, simply because this system remains good for running fuzzy searches, that is to say similarity searches, where a certain lack of precision is necessary (that can be especially useful when assessing inventive step, see as an example EPO jurisprudence T0164/83).

It should be said that other fragmentation code systems were developed at that time, especially the GREMAS code designed by the IDC (International Documentation in Chemistry) consortium established by Hoechst, Bayer and BASF.

Two rival consortia

Another major player in the chemical information area was on the lookout for a topological search system: CAS, which had introduced in 1965 the CAS Chemical Registry System. CAS decided to team up with Télésystèmes Questel, a subsidiary of the France Telecom group, and both partners succeeded in launching in 1981 the first topological search system – the so-called DARC system –, permitting to store and search specific compounds (compounds without any variability).

This probably partly put Derwent under pressure: in other to stay in the game, it appeared necessary to move forward from the fragmentation codes to something more innovative matching the evolving need of the customers.

At the same time in France, in the late 70s, it appeared key to some political decision makers to develop an independent French patent database vendor in order to support the growth of the national industry, especially the pharmaceutical sector, quite strong in France at that time; hence the project to cope with the Markush issue. This idea was promoted by one of the influential persons at the French ministry of research and higher education, Jacques Michel, who later became CEO of Questel. Later on, J.Michel joined the EPO where he took over the position of vice-president. It seemed logic to allocate the task of building the new Markush patent database to Questel – a subsidiary of a French public company – and to INPI, the French patent office, especially as INPI was already involved with Questel in the patent database area. INPI, a true pioneer in the patent information area in this period, was indeed the first office in the world to have launched (in the mid-80s) two online patent databases: INPI1 covering the French patent applications, and INPI2 covering the EP ones.

Competitive challenge on one side of the channel, political decision on the other, the same appetite for cutting-edge projects on both sides, that's how INPI, Questel and Derwent finally decided in 1983 to combine their resources to build, from the already existing DARC system, the Markush DARC one.

The cliff was very high: imagine, making searchable compounds potentially covering thousands of combinations.

At the same time, the challenge was quite thrilling, as we were alone in the world to work on such a topic ….. nearly alone actually, as CAS on the other side of the ocean, was in the same period trying to design its own Markush system, working from the Messenger software, launched some years before, and already able to cope with specific compounds.

This story was in fact due to be characterized in the next 20 years by a transatlantic competition between two groups of people:

From left to right: Jacques Michel, Jacques-Emile Dubois, Pierre Buffet and Serge Chambaud.

On the European side, we had Jacques-Emile Dubois (1920-2005), physicist and chemist, inventor in 1965 of the DARC topological system at the Paris VII university; Pierre Buffet, Engineer in Informatics, cofounder of Questel, major contributor to the cooperation between Derwent, INPI and Questel aimed at developing the Markush service; at INPI, Irène Savignon, head of the database department, and behind the INPI involvement in the patent database area. When she retired, Irene was replaced by Serge Chambaud, previously member of the Dubois's team, head of the INPI documentation department until 2005. And as I said, Jacques Michel, at the state secretary for research in the late 70s, one of those behind the foundation of Questel, managing director of Questel in the mid-80s.

From left to right: Michael F. Lynch, John Barnard, John Holliday and Robert Massie

On the US side, we had Michael F. Lynch, professor at the Department of Information Studies at the University of Sheffield and pilot of the developments which produced the structural search algorithms used on STN. John Holliday, in the Lynch team, Research Manager in the Chemoinformatics Research Group at the University of Sheffield. John Barnard, in the Lynch team as well, later Scientific Director at Digital Chemistry Ltd. And Robert Massie, President of CAS between 1992 and 2014, who contributed to making CAS one of the key players in the patent database area.

The contract between Questel, INPI and Derwent had been signed 3 years before I joined the project in 1986. This contract is probably one the very last projects in which Monty involved Derwent.

Early days of drudgery

I started at INPI as an "indexer". An "indexer" (or a "coder") was a guy equipped with pens and paper sheets, reading patents and redrawing the chemical formulas contained in them according to the complex Markush DARC representation rules; the resulting paper sheets were then transferred to people using a mouse and a keyboard connected to a computer to input the drawings using specific drawing interfaces. The resulting output files were then transferred on magnetic tapes to Questel in order to be loaded in the Markush database. I remember that we had to index a minimum of 10 patents per week. We were a little bit the seamstresses of the early 80s. When I joined INPI, the job was assigned to patent examiners on a part-time basis. A team of 10 staffs totally dedicated to that task was then created in 1987. Later on, it was also outsourced to 3 subcontractors. Overall, in the early 90s we had about 50 staff working on the project, including the administrative officers and managers.

It should be borne in mind that chemical formulae were not provided by the applicants in electronic machine-readable format (and they are still not). Since the mid-90s, applicants must provide peptide or DNA sequences, as well as text, in electronic machine-readable format, which permits the patent offices to smoothly load that type of data in free

accessible databases. At the time I write this text – June 2020 – providing chemical formula in such a format is not yet compulsory. Hence the need to have teams of people reading the patents and manually producing an output format suitable to be loaded in a database. Of course, progress has been made since the 80s. For example, OCR software have been developed which enable to automatically digitize the chemical formulas, which makes the need for human indexers lower. But this does not apply to Markush formula.

Why had we to enforce complex representation rules? simply because the Markush syntax allowed in a patent claim is so wide, so free, that it could not fit with the possibilities of a digital representation. The designers of the Markush DARC system had to imagine a specific syntax -- of course narrower than the original one in the patent -- to make it possible to store the information in a computerized format. The challenge of the indexer for each claim was to have it fit within the Markush DARC syntax, and pretty often the "foot was larger than the shoe". As a result, the representation obtained in the database record did not exactly match the original. Sometimes it was somewhat broader, and sometimes a bit narrower. The quality of the copy made the system good enough, however, for searching.

Needless to say, the job of indexer rapidly became boring for me. In 1989, for various reasons, the Markush project manager left INPI. We faced the big challenge to find somebody to replace her. INPI is now a modern patent office installed in a brand new environment-friendly building in the Defense district – the "Paris city" area – and recently hired dozens of young and dynamic examiners provided with AI-based tools to perform their daily work. But you should imagine that in 1986 INPI was a traditional administration, based in a 19th century pastiche of middle-age convent, with its vaulted chapel, traversed by conscientious civil servants, groups of whom (wearing dark over coat and felt hat) used to punctually wait at five to 5 pm before the punch clock in order to start their journey back home at exactly 5 pm. In this landscape, the Markush team looked a little bit like a UFO. And the Markush project manager was a true ET, often travelling to London to discuss with Derwent before flying to the US to participate to some ACS meetings, working at least 50% of his time in English. In addition to this, those overconfident guys requested by the end of the 80s a LAN and an e-mail facility to be installed at INPI! nobody at INPI had this style of life nor this type of claim.

Working in English seemed an unachievable goal for everybody, and I inherited the position of technical manager, which dramatically changed my life. Locked in the backstage until then, I could fully enter in the project.

Technical meetings pool expertise

From that year on, I started to participate to all the Questel-INPI-Derwent technical meetings, partly in London at the Derwent premises, first in the Holborn district at Rochdale House, then 14 Great Queen Street, partly at the Questel head office in Paris or Nanterre; never at INPI, would you ask? no, nearly never at INPI, as hosting such a meeting meant inviting partners at restaurant, which, due to the bureaucratic procedures by which we had to abide, was a quite tedious obstacle course.

The aim of that meetings were to manage the project, for example to decide about the future capabilities to be introduced in the interface, to plan the tests, discuss their results with Questel, prioritize the corrections to be done, organize the communication, the training, the beta-tests with the users. We had to design the structure of the database, the loading software to feed the database, the search capabilities and the search interface incorporating such capabilities, the off-line indexing software used by our subcontractors to input the chemical indexing ….; we had also to make choices about indexing rules: how to index tautomeric moieties, how to cover definition like "R1 can be any carbon chain", …. Although INPI and Derwent had made the decision to produce their own Markush file, MPHARM for INPI and WPIM for Derwent, the indexing rules enforced by both organizations were essentially the same. In 1990 the project was able to take advantage of EC funding, as a result of which we also had to manage relations with the EC staff.

I remember my first meeting at Rochdale House. I found myself in a room with Cathy Shenton, Peter Norton, Gaz Cross, Tony Ferns from Derwent; Bill Town and Dave Proctor from Hampden; Pierre Buffet and Dominique Renaud (Pierre's technical arm, later joined by another key technical arm: Pierre Beni chou) from Questel; and Catherine Pagis,

marketing manager for the Markush DARC project at INPI. I was impressed and …. lost: my very scholastic English prevented me from following 80% of the discussion, and Catherine had to wake me up when somebody asked me a question: I had not realized that I had been addressed. Cathy Shenton had a US flag on her sweatshirt and I understood quickly that she had a strong character, she knew what she wanted. Peter Norton was a true scientist, modest but really knowledgeable.

Peter Norton (Derwent), expert in the indexing of « nasties »

One of our problems was the indexing of the so-called nasties: a nasty was a quite complex Markush claim, having a high number of alternatives with deep levels of nesting (R groups containing R groups containing R groups ….). The problem was to index those formula: we couldn't afford to spend ages on the work. It was therefore necessary to find ways to speed-up the indexing while covering the information reasonably well. Peter later produced, within the framework of the EC funded project, a report on the indexing of nasties.

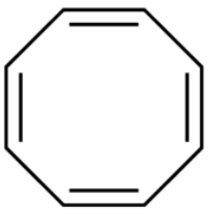
If Cathy and Peter left the ship relatively quickly, I was destined to maintain long relations with Gaz and Tony, who have really been two long-lasting pillars of the Markush trip along the nearly 20 years I have spent involved in it. They were very professional, quite discreet, with a typical British sense of phlegm which I did appreciate. Bill Town was really a smart, elegant, easy-going and very diplomatic person. He could speak both French and Italian. Some years after that meeting, he left Hampden and joined Derwent, leaving the management of Hampden to Peter Nichols. Pierre Buffet really impressed me, he was a really smart person, having a high professionalism, a quite acute sense of humor and an especially impressive mastery of the English language. If I could, after some months of meeting with Derwent, manage to cope with English, that's especially thanks to him. I used to grab every new word I heard from his mouth, one of which – I still remember that – was "definitely", which he used quite often. He retired about 8 years ago and as far as I know he enjoys now a peaceful life in his house in Provence where he spends much time growing his rose trees. Tony, Gaz, Pierre, Dominique and I took part in nearly all the meetings we had. We were joined in 1990 by Bernard Marx, a specialist in the documentation area, recruited by Irene Savignon. When travelling to London, we used to meet around 5 am in the center of Paris before travelling together to CDG airport. In this pre-Eurostar era, getting from Paris to London was a real trip.

We were all aware that this work was leading-edge for its time and it was …. definitely exciting to participate in it.

It was during one of these meetings at Rochdale House, may be in 1989, that I had the opportunity to be introduced, with other French colleagues, to Monty. Later on, in 2000, I was at the Derwent subscriber meeting when he received the IPI award.

A fist-fight nearly breaks out

Of course, the life in the Markush family wasn't always a long and quiet river. I remember especially an occasion where both "indexing" parties (Derwent and INPI) had a quite serious argument. Our two "chief chemists" nearly came to blows, under the totally puzzled, and dismayed, eyes of Pierre Buffet.



The object of the 100 minute battle between the Britons and the Gallics in 1991

The topic of the disagreement related to what we called "translation". CAS had implemented on MARPAT the capability to search for families of chemical moieties by simply using a 2- or 3-character symbol: Ak for example permits to search for any carbon chain. Markush DARC did not have this quite useful tool, and our users had to use quite heavy variable groups having long lists of alternatives (e.g. R = -methyl; -ethyl; -propyl; -butyl and so one) to carry out what was so simple on Marat. As a result, we were asked by our customers to implement an equivalent function. In order to tackle the issue, we imagined the Superatom system, 2 of which were dedicated to carbon rings: ARY, to retrieve all aromatic rings; CYC, to retrieve non aromatic rings. The question was: is the cyclooctatetraene ring aromatic or not? Should it match ARY or CYC as part of the "translation" capability? Many arguments were exchanged (at the end,

thrown to the other party's face I should say). Finally, considering the intransigence of the fighters, and the impossibility to call a UN peacekeeper, our chief-diplomat – Pierre Buffet I mean – decided that for this specific case 2 versions of the software would be developed, one for INPI and one for Derwent. Some days later, I ran a search on our Markush files, and realized that they had only 0.01% of such cyclooctatetraene rings. Huge tension for a tiny issue.

We finally managed to launch the so-called Translation capability, which permitted us to catch up Marat, together with another one, the possibility to define variable positions of attachment, which permitted us to be one step ahead of our competitor.

Indexing – a greater controversy

In the course of their relation, INPI and Derwent had another area of disagreement the impact of which was more significant. It was related to the indexing policy. In the above figure 1 about specifics and prophetics, I explained that when building a claim an applicant usually consolidates in the R groups the alternatives corresponding to the various examples, adding for each category of alternatives a generic term representing the generalization of said category. For example, if a group has the methyl, ethyl, propyl and butyl groups (corresponding to the synthesized examples), the generic term "carbon chain having from 1 to 6 carbon atoms" can be added. It can sometimes happen that a generic definition is present in the claim without corresponding specific term in the examples. For instance, the applicant includes the value "alkenyl" in a R1 group whereas no example having a specific value of alkenyl for R1 – e.g. –CH=CH2 – is present among the specific examples described in the patent specification.

Other example: a claim specifies that an R group can be connected on the position 2, 3 and 4 of a pyridine ring, while the listed examples have only compounds where this R group is connected on the position 3.

As I said above, Markush claims were often quite complex to index, and in order to simplify the burden of the indexing INPI decided to index only those generic terms having a related specific one exemplified in the patent specification. Alkyl was indexed only if methyl (or any other specific saturated carbon chain) was present. In other words, we had in mind a Markush file closer to the examples.

Derwent disagreed with that, as it prevented fully covering the scope of the claims, making our database quite unreliable to run an FTO search. Derwent was definitely correct in this, and the position of Derwent was actually in line with the needs of our customers, especially BASF. For my part, I had no idea what the decision should be, simply because my knowledge in the field of IP was very poor at that time …. which is amazing for an employee of a patent office. I was more expert at using the Markush DARC interface than in dealing with the tenets of the Intellectual Property Code. I guess that other INPI staff, more involved than I in the indexing side, had a position simply reflecting the need of a patent examiner: such a guy never has to run an FTO search, since FTO is not the problem of a patent office. During the patent examination procedure, one simply needs to check novelty and the inventive step, and a database covering examples is enough to check novelty. I discovered later – when I had the opportunity to bury myself in the subtleties of the IP rules – that examples only are not enough to check the inventive step; and that the primary interest of a Markush file is the fact that it covers prophetics, in order to make it suitable to run FTO. To my mind, the INPI position was therefore not relevant.

What does MMS stand for?

Until 1998 INPI and Derwent, although taking advantage of the same search software – Markush DARC – they had jointly designed with Questel, produced their own file, MPHARM for INPI and WPIM for Derwent. This represented a costly duplicate of work. In 1998 both database producers reached an agreement to produce a single file named MMS. MMS officially stood for Merged Markush Service*. But Pierre Buffet had another idea about that. Pierre said "when I was 20, MMS meant* Matin, Midi et Soir *(Morning, Noon and evening). At 40 that was only* Mardi, Mercredi et Samedi *(Tuesday, Wednesday and Saturday). At 60 it became March, May and September. And now, that is* Mes Meilleurs Souvenirs *(My best Memories).*

May be a good example of the French sense of poetry.

However, as long as INPI and Derwent carried on indexing their own respective database (MPHARM and WPIM), they used two different indexing policies. Later, when we decided to combine our resources by merging both files – which produced the Merged Markush Service, from 1998 – INPI made the wise decision to adopt the Derwent indexing rules.

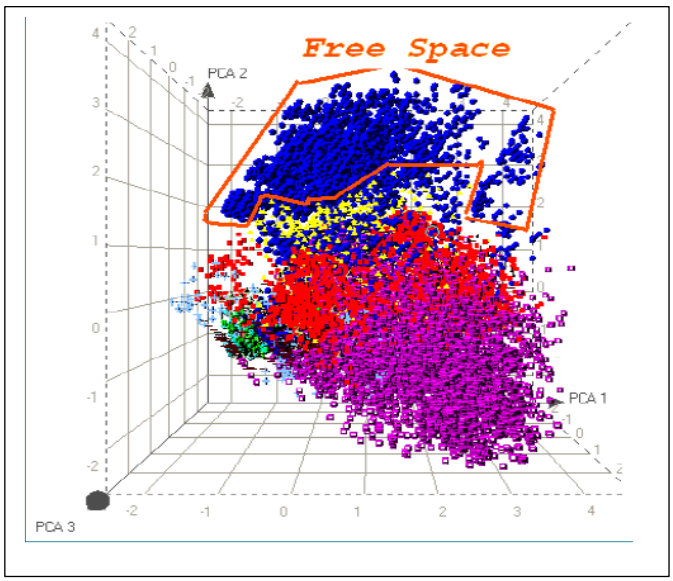
The manager of IP information at INPI in those years – Serge Chambaud – had a state of mind which was really in line with Monty's: always on the lookout for new opportunities to innovate. Irene Savignon and Serge Chambaud's basic assumption was that a patent office is at the service of innovators and that in order to fulfill its mission it must be itself innovative. In a context where the big patent offices play an increasing role, a national patent office is sentenced to fade away if it doesn't play the card of innovation. In the mid-90s Serge launched a system generating the IPC codes from a technical text used as an input, which permitted to greatly simplify the usage of the IPC for the patent searchers. Some years later, he involved INPI in a project related to a figurative trademark search system based on an image recognition software. In the years 2001-2003 the Markush DARC project generated a spin-off benefit which Monty would have loved.

Pfizer sets a challenge

The Pfizer people in Groton (MA) were willing to implement a strong analysis tool enabling one to process large numbers of claims/examples with similarity ranking/overlap analysis capabilities. The purpose of that tool was to analyze Markush claims, identify overlaps between claims and white spaces. Pfizer already had imagined a process having two steps: first an enumeration, aimed at generating all the prophetics theoretically embedded in the claim to be analyzed; then the generation of a graphic representation of the enumerated compounds as a cloud of dots. Pfizer had managed to cope with the second step by implementing an algorithm using a PCA (Principal Component Analysis) based on atom-pairs. This was a "Pfizer Blackbox". Spotfire was used as a visualization software. The first, enumeration, step remained a problem for Pfizer, and they decided to organize a kind of call for tender, gathering all the organizations involved in chemical information in Groton (CAS, Derwent, INPI, ….) to listen to their propositions. And finally the Markush team at INPI was selected to face the challenge.

Our strong experience in the Markush indexing area convinced Pfizer that we were the right people to work with.

We had some problems to fix in the course of implementing our enumeration software. We had for example to cope with those generic values such as "5-6 membered heterocyclic ring system having one N atom and another heterom taken from N, O and S". How to generate all the specific values matching this generic definition! As a solution we produced libraries for each generic definition (Superatom libraries); of course, these libraries were not 100% comprehensive: it's impossible to be comprehensive as the potential number of values falling into the scope of such generic definitions has no limit. Another issue came from the Markush DARC syntax which, as I said above, prevented the MMS records to exactly reflect the content of the related Markush claim in the patent. As a solution we had to devise extended indexing rules and documents to be treated had to be 100% reindexed. Another problem came from the fact that an enumeration can generate hundreds of thousands of individual compounds. We had to put in place a random enumeration allowing the user to define a % of enumeration.



Our work resulted in a prototype producing the kind of diagram alongside. In this diagram, we compare a new Markush claim (the compound which we invented) to prior art Markush. Blue dots correspond to our invention; the others to the prior art Markush. The free space is clearly identified by the red lines. Such a tool could be used for two purposes:

- by companies about to apply for a patent, in order to check the possible presence of "empty areas" in their own Markush; as an empty area in your Markush means a possible entry for your competitor.

- by companies willing to check the Markush of their competitors, in order to identify possible loopholes in these Markush.

We used the Viagra patent to test the prototype, and I remember that Anton Fliri – the project leader on the Pfizer side – looking at the above diagram said “Had we had that product when we filed the Viagra patent, we would have avoided the Cialis”. The main limit of this project, was the requirement to reindex documents in order to obtain an accurate analysis; commercially, it was hardly affordable.

I worked on this Markush ship until 2005, when our CEO decided to withdraw INPI from the battle, for financial reasons.

Reflecting on the experience

Throughout the 20 years I spent on it, this project was a great experience for me personally. Leaving the narrow horizon of one’s native country, working at an international level, discovering new ways of thinking, generate great benefits. It was permitted for me to use, speak, understand, and write English, and to work with many people from different countries -- quite an exciting side of the adventure. I could in particular work for some 15 years with people from the UK, especially with the Derwent staff, and together we could reach a quite fantastic result. This was a thrilling experience, also from a personal point of view. French people should never forget that London is the city where, somewhere in the 40s, France could be brought back to life. The UK is a key component of Europe. Today, in the current quite uncertain world, I strongly believe that our future will be easier if it is common.

I could share a part of my professional life with some wonderful and inspiring colleagues, top experts dedicated to excellence. I already mentioned many. I should add Brian Stockdale from Derwent, a top-trainer I would meet in the course of our visits in Germany and in the US. I also absolutely need to mention Mike O’Hara, president of Questel Inc. and then our representative in the US. Mike was a great professional, a really subtle and spiritual person, and working with him was a great pleasure. I remember one of our last telephone conversations, spent on working on the MMS user manual during nearly 3 hours. After picking up the receiver to answer my call, Mike said “oh, the TV is announcing a problem in New-York, an air-plane crashed in a tower”. That’s probably an accident I said, imagining a small inexperienced tourist airplane. 3 hours later our conversation was sharply interrupted. That was on September 11.

I also should mention our customers. Most of them were very knowledgeable, and on many occasions the courses we gave were opportunities for us to learn also. This was especially the case with German users, some of which had been themselves involved in the design of in-house patent databases. I shall mention Winfried Detlefsen and Werner Jünemann and the team at BASF in Ludwigshafen, Gesche Berger and her team at Hoechst in Frankfurt and also Gerhard Fischer and his group at Syngenta in Basel. Nor can I forget our users in the UK.

Although operating in a quite narrow and specific niche, above the clouds, some events reminded us that we kept connected to the ups and downs of real life. Which happened in 2005 when we learnt that Kitty Silva, a patent searcher at GSK, passed away, victim of the tsunami in Sri Lanka in December 2004. Kitty was a very gentle person, and as far as my learning of English is concerned, I grabbed “I have been tied up with” from her, so each time I use this expression, I can’t avoid thinking to her. At INPI, Irene Savignon, in charge of database policy in the 80s, was the main initiator of the project, and I will remain always grateful to her for having involved me in it. She was – and probably still is, she was born in 1928 – a quite passionate and very inspiring person. I recently discovered the hardships she endured in her early life. That’s probably what gave her a great sense of humanity and the strong determination she has shown on many occasions, especially in her professional life.

High cost, low income, but significant benefit

As I said above, the INPI head decided in 2004 to pull out, which decision took effect in September 2005. It had become less and less sustainable to invest money in a product which was a technical success but could only attract the patent information departments of the top 50 chemical and pharmaceutical companies in the world, thus generating little income. Moreover, 90% of our customers were based outside France (50% in the US, about 20% in Germany, and the rest in the UK, Scandinavia, Belgium and Netherlands). For our regulatory authorities it became hard to understand that a French public administration could invest money in order to meet the needs of non-French businesses.

I would however balance that approach by saying that MMS had many beneficial impacts for INPI. In term of image and visibility at the international level first: we were recognized as a truly innovative office in the IP community. We could work on a peer-to-peer basis with all the major players in the chemical information area at that time and could establish long-lasting relations with Questel, Derwent and also CAS, and could commence them with the Chinese Patent Office. The doors of the European and US chemical and pharmaceutical industries were open to us, we did not have to find our customers, they usually went to us; we were regularly invited by the PDG and could participate to the Derwent subscriber Meetings.

I especially remember a meeting we organized at the St. Francis hotel in San Francisco where we were praised by the US patent information community for our contribution in the chemical information area. That was quite unique for a patent office, and only the "big ones" of that time (EPO especially) could play at the same level.

Secondly, MMS brought INPI a lot in term of know-how transfer. As I noted before, in the early 90s INPI remained a little bit traditional, and the fact of working with private foreign companies permitted us to significantly progress in some key areas: the first LAN and e-mail facilities were introduced at INPI as a consequence of our project, simply because we realized that Derwent and CAS already had invested in such tools, and that in order to work with them we simply needed to implement them as well.

Without the INPI funding, it became hard for Questel to maintain and enhance the product. Customers probably started shifting to Marat on STN, which trend was accelerated by the fact that in 2015 the Derwent part of MMS was loaded on STN, making that host the "one-stop shop" for searching both the specific compounds (in the Registry file) and the Markush ones (in Marat and in the Derwent Markush file). Finally, Questel made the decision to discontinue MMS in September 2020. That results in an overwhelming position for CAS, which is not necessarily good for customers. That takes us back to the possible "original sin" of the patent offices: not having made the electronic filing of chemical information mandatory.

Overall, that project was for sure one of the most inspiring parts of my career, a little bit *Mes Meilleurs Souvenirs*, as Pierre Buffet would say.

I shall express my deep gratitude to Peter Hyams for the thorough revision he made on this text; the 19 years I have spent dealing with Derwent could not manage to make me a true Oxford citizen! Many thanks also to him for hosting my paper on this quite inspiring web site.

Philippe Borne, INPI, member of the CFIB and PIUG

I started at INPI in 1986, and worked for 19 years on the Markush DARC project mainly as a technical/training manager. In 2006, I took over the position of deputy manager of our patent searching group, in charge of offering on a commercial basis patent searches to external clients. In 2008 I stayed at INPI but transitioned to a totally new job, as I was appointed as INPI representative in Burgundy. As such I had to spread the IP culture throughout the local network of innovators, especially SMEs. Top goal: to convince them that IP is key for their competitiveness. In 2014 I moved to Strasbourg where I do the same job together with the 9 staff I manage, in an environment characterized by its proximity with Germany (and a German-type state of mind), a pretty high concentration of businesses (with a significant number of subsidiaries of German and Swiss companies), a high level of scientific research (Strasbourg University produced 4 Nobel Prizes) and a good level of IP awareness which probably results from the fact that Strasbourg hosts the CEIPI (the Centre for International Intellectual Property Studies, which trains 100% of the French patent attorneys). And I shouldn't forget the Christkindelsmärik (the Strasbourg Christmas Market, held annually since 1570).

The views and opinions expressed in this article are those of the author and do not necessarily reflect the policy or position of his employer or any organisation to which he belongs.