

## 帕特·塞林格访谈录

本专访介绍了System R的成功之道，如何与INGRES和QBE团队合作，如何进行技术转化，如何进行人员管理，如何同时领导3000个人，以及更多。

玛丽安 温丝莱

欢迎来到本期SIGMOD Record数据库杰出人物的系列访谈。我是玛丽安-温丝莱，现在我们正在圣地亚哥，这里是2003年SIGMOD 和 PODS会议的举办地。现在我正和帕特 塞林格在一起，她是IBM数据管理架构和技术副总裁。帕特是数据库领域最有影响力论文之一的主要作者：她在1979年SIGMOD上的论文-“关系数据库管理系统中的数据访问路径选择”，为今天的主流数据库系统查询优化器奠定了理论基础。帕特同时致力于将IBM的创新研究应用到商用产品中。帕特是IBM院士，ACM SIGMOD创新奖获得者，是国家工程院成员。她在耶鲁大学获得博士学位，让我们欢迎帕特！

**玛丽安：**帕特，System R和INGRES向反对者们证明了构建一个性能良好的数据库管理系统的可能性。作为早期System R团队的一份子，你见证了一个新兴行业的诞生。构建System R过程中你最大的惊喜是什么？在这个过程中你经历过哪些特别正确或特别错误的事情？

**帕特：**我来到IBM的时候System R刚刚起步。已经有人做了一些系统底层工作，但是上层工作刚刚开始。当时我具有操作系统和编程语言的背景，但是对数据库系统一无所知。所以他们给了我一本Chris Date的书并对我说“读读这本书，读完以后开始和我们交流”。令我非常惊讶的是我的编程语言和操作系统背景和数据库系统有那么多匹配的地方，而数据库领域超越了编译和设备驱动技术以及内存和I/O管理。数据库充分利用了其他领域的技术，然后把他们构成一个有机的整体。所以就我个人而言，意识到我确实能对System R有所贡献令我非常惊喜。

**玛丽安：**那么对整个项目来说，有什么特别正确或者特别错误的事情发生吗？

**帕特：**我想我们有一个非常了不起的团队。那时我们这个项目有30多个人，他们当中有人在这里读博士后，有人在项目中期加入。相对于普通的学术环境，在那里你一个人在为博士学位奋斗，我们团队中的协作精神以及家庭般的温暖非常浓厚。这种氛围对我来说是耳目一新的，而且我一直在IBM工作的很大原因是我非常喜欢这种相互协作的工作方式。

**玛丽安：**现在回顾System R的发展历程，您觉得它成功的关键因素是什么？

**帕特：**我认为很多公司采用 SQL 语言并实现它这个事实使得 SQL 成为最广泛的数据库接口。因此这也是 SQL 成功的一个重要因素。另一个重要因素是我们事先与很多公司签约使用该系统，在我们开发的过程中就在使用。波音公司是我们的早期客户之一，他们说，“System R 是一个了不起的系统，但是现在还在二月，我们已经用完了一年所预算的计算时间来进行如此多的数据库查询”。然后我说，“我们抱歉，真的很抱歉！”他们说，“不，不要抱歉，System R 非常棒。我们现在可以问一些在以前永远不会得到答案的问题，我们现在可以做很多以前不能做的事情。”因此我认为 System R 给用户带来的应用开发效率的提高使他成为一个巨大的赢家。事实上我们的客户告诉 IBM，从 IBM 得到的支持对于他们的改变非常重要。

**玛丽安：**一件很有意思的事情是，我问过Bruce Lindsay和刚才问你一样的问题，所以当他的面谈发表以后，读者们将会看到一个关于“SystemR为什么如此成功和有影响力”这个问题的不同解释。但是我不会泄露任何秘密。我不会告诉读者Bruce究竟说了什么。读者们必须等到Bruce

的面谈发表以后才会知道他的答案。

在开发System R之初，要判断他给世界带来的影响是好还是坏是不可能的。现在来看System R的发展过程，如果给你一次机会重新开发System R，你会做什么不同的事情？

**帕特：**回想起来，我没有足够的经验作为一个系统设计者和开发者来按照IBM内部人员预想的那样实现System R，并且进行产品化的复制和销售。因此设计控制模块，设计接口和API编程语言——我们只是把它们按照我们觉得可以工作的方式组合起来。错误处理模块，用户输入变量和参数的方式——这些都没有足够的思考来设计。它们组合起来仅仅是能工作，我们也没有考虑过如何设计使得系统具有扩展性。我们没有考虑过用户可能要表达的一千种其他的数据类型。因此我想如果重新开发的话，我将会考虑得更全面，这样System R的应用程序编程接口将会更自然的适合使用它的编程语言。

**玛丽安：**今天INGRES和System R在数据库历史上的地位都是非常安全的。不过在这两个系统设计之初，两个团队在同一个时间以同样的目标解决同样的问题，这不可避免的带来竞争。这种竞争的形式是什么样的？

**帕特：**我认为这是一种良性竞争，我们在解决一个设计空间的不同方面：不同的展示视图，不同的查询优化方法等等，这对数据库领域的发展非常有意义。我们要探索比原来做得任何个项目都要多的解决方法。所以我认为当时的工作对工业界的发展是非常有益的，它推动我们前进。

**玛丽安：**事后看来，如果你能回到那些日子里，会如何处理这种情况？我问这个问题是因为今天和未来，其他的研究者会发现他们可能会处在同样的情况之下，而您的意见对他们会有帮助。

**帕特：**我觉得现在有一些平台可以供我们进行讨论以做出更好的工作，例如 SIGMOD——讨论一些技术的优势和劣势。如果我们能有一些更加开放的讨论，在一个领域内供大家共享，我想这会比两个团队各自做自己的工作而不进行交流更好。

**玛丽安：**在IBM内部，QBE和System R被视为竞争对手。现在看来，有没有更好的方法处理这种情况？

**帕特：**我不认为我们的处理方式存在问题。我们有两个团队从不同的方面解决相似的问题，可能有一些人认为存在着竞争。我事后认为，QBE的工作专注于用户界面，以及用户如何定制查询，它们使得容易的工作变得更加容易。SQL和System R更面向“重型”的工作，我们更专注于深层技术，扩展性，性能等。这不是QBE的工作重点。我想如果当时有人更早的使之合理化，那么就不会有一些冲突的感情和看法出现。

**玛丽安：**因此人们没有完全理解这是两个贡献，一个在用户界面层次而另一个——

**帕特：**这里从来就不存在竞争。

**玛丽安：**那OK Corral中的Shoot Out呢？（有兴趣的读者可以从网络上20年前System R的文档中看到Shoot Out）

**帕特：**你是对的。我太老了！那是一个性能比赛。

**玛丽安：**似乎不是那么正确——

**帕特：**鉴于两个系统的侧重点不同，单纯比较它们的性能似乎不太恰当。我们一直做的工作是将QBE的前端和System R的后台结合成一个系统。

**玛丽安：**是的，单从性能方面来说，System R肯定要优于QBE。因此我想这种结合应该实际上没有发生——

**帕特：**其实不是。两者的结合正是我们在QMF产品中实现的。因此这种结合我们已经实现，只不过是产品实验室，而不是研究实验室。

**玛丽安：**喔，了解。IBM一直成功的将研究成果转化成产品。这些年来你一直负责这项工作。成功转化技术的秘诀是什么？

**帕特：**在我创建IBM数据库技术研究院（IBM研究组织，致力于数据库技术的转化）时，我做了两件特殊的事情。我找到IBM数据库领域的每一位项目经理，坐下来和他们交谈至少一个小时。我和他们讨论他们的目标，他们的问题，并解释我希望这个团队做些什么以及我将如何与这个团队交流。这就消除了笼罩大家内心的“我将和你竞争”的恐惧。我们做的第二件事是在会议上共同讨论我们要进行的工作，开发人员面临的重要问题，未来一年的工作议程。我们研究团队有的人常驻在开发实验室，和开发人员一起参与到讨论与工作中。从而使开发人员证明了他们的“价值”。在做这些事情的时候，有一个“相互尊重”的理念在其中。因此这种团队合作的工作方式就是并肩作战。

**玛丽安：**年度会议就是开发产品的人告诉研发人员他们的问题吗？

**帕特：**嗯，这是为期一周的会议。一周开始研发人员告诉我们，“这是我的发展计划；所有的事情我觉得都在可控范围内；这些事情我认为需要的时间或更多的发明，才能让产品正常发布。”然后这一周接下来我们回到数据库技术研究所，谈到“过去的一年所做的事情”，在这一周的周末，坐下来，我们做了一个排名列表，哪些事情是在接下来的一年中我们想投入的，这不论我们自己还是IBM都是很重要的。这样列表中对于多产品组中有价值的东西会更多。

**玛丽安：**因此，这是产品团队的需求的需求和你的具体工作之间的一个匹配过程？

**帕特：**是的。

**玛丽安：**所以并不是你们看到他们的需求，然后你们去做研究以满足他们的需要？

**帕特：**我们整个数据库技术研究所的研究人员整周都在听演讲，无论是茶歇还是晚餐，团队中每个成员都会进行沟通，相互了解。我们会讨论，“那么，这种方法解决您的问题怎么样？”，“我用这个技巧来解决你的问题”，和“如果以这种方式解决这个问题，怎么样？”在周末，人们都会对与工作相关一些事情有比较好的想法。

**玛丽安：**在数据库技术研究所的人，他们是全职员工或者是属于研究所的研究人员？

**帕特：**数据库技术研究所是一个俱乐部。他们都不是为我个人而工作的，完成工作是所有人的心声。所以这是一种主人翁的精神，我们发现使工作变得更有意义的一种方法，这会使顾客对于我们的工作成果会更加满意。大部分项目都会既有研究团队的人，也有开发的人。

**玛丽安：**还有一件事：你提到的消除竞争方面。你能解释竞争的出现吗？

**帕特：**在有些其他的公司可能会出现研究部门和开发部门合作的不是很好，而研究部门的人在论文发表或者系统开发后就不再进行更多的工作。他们在这一点上的态度是，“我这样做，你难道觉得不好？”而自动生成的响应，“我可能做的更好，我有不同的设计，你不明白我的意思，你不了解我的客户。”如果情况是这样，合作起来会很难。如果失效的话，那么在感情上就会受到影响。所以，如果我让你从一开始就参与，你觉得自己是在帮助塑造和引导这项工作，并告诉我重要的事情，那么情况自然会好很多。

**玛丽安：**是什么促使你加入工业界并留在IBM，而不是从事学术研究或者去其他的公司工作？

**帕特：**我决定在加州工作是因为我曾在波士顿住过八年，那里寒冷泥泞而且冰天雪地，但是加州就好很多。因此我对自己说我要在加州找一份工作。我曾经拿到加州大学圣地亚哥分校的offer，以及IBM圣何塞的offer。我观察了将要和我一起工作的人以及我将从事的工作类型。在学校从事助研工作时，我参加团队开发工作并且开发了一个编译器，我很喜欢这种团队工作。在学校的那一段时间是一种每个教授负责自己工作的机制。是的，作为一名教授你有自己的研究生，他们帮助你工作，你和你的学生形成一个团队，但是那不等同于你拥有十五个人总是待在那里和你一起工作。因此这很吸引我，现在也是。

**玛丽安：**你没有像很多参与过System R工作的人那样跳槽，是什么让你留在IBM？

**帕特：**我喜欢这里的人，顶尖的人。这应该是主要的因素。而且我喜欢IBM在工业界的这种影响，而且我看到更多的潜力。

**玛丽安：**我知道您在IBM从事导师的工作。如何使一个刚毕业的博士尽快转换成研究员，并使他的工作产生有意义的影响？

**帕特：**我认为对于一个新毕业的博士来说最困难的转变是从个人研究的思维模式，转换成工业界研究团队的一部分。对于有些人来说这种转变非常困难，因为他们没有接受过团队工作的训练。他们不理解“只有当不破坏你的代码时我才能修改代码”的软件开发守则。在研究部门接触到的项目一般是由五到十个人构成。但是如果你参与到开发团队中，例如数据库技术研究院的项目，那么你将会有200到600个人一起为全球实体开发软件。

**玛丽安：**那么你如何帮助他们进行转变呢？

**帕特：**我们采用Hamid Pirahesh提出的一个概念，叫做“移民项目”。在项目中每个人只知道系统的一小部分，在导师的帮助下完成整个项目。你会适应这种环境下的软件开发模式和设计模式，这种模式比一个有经验的研究员或者开发人员的随心所欲工作模式要稍微结构化一些。

**玛丽安：**你曾经管理过Jim Gray和Bruce Lindsay。Jim建议我问问你管理像他和Bruce这种比较麻烦的人员的秘诀。你是怎么做到的？

**帕特：**是的，我不仅管理他们俩还有Franco Putzolu，他现在是Oracle的院士。他们三个在同一个时期在我手下工作，是我作为经理的第一批员工。他们热情，有激情，思想活跃而且不轻易同意别人的看法。我向秘诀就是我真的非常喜欢他们，并且尊重他们。你知道，你越是更好的照顾他们，他们就会更努力的工作。

**玛丽安：**是的，可能是这样，可是他们三个中有个人告诉我，你可以让他们每个人做任何你像要他们做的事。有什么秘诀呢？

**帕特：**就是因为我喜欢他们。

**玛丽安：**喜欢他们是指...

**帕特：**我很关心他们，我努力去做一个有说服力的人，一个共识的缔造者，而不是一个有独特风格的管理者。这也是数据库技术研究中心成功的原因。我把大家召集在一起，让大家为了一个共同的目标而工作。

**玛丽安：**但是现在你是领导几百个人的副总，不可能关心他们每一个人。那么现在你又如何处理呢？

**帕特：**我现在基本上成为IBM数据库研发中心3000个人的“团队妈妈”，他们可以随时给我打电话。他们有我的家庭电话，我的手机号码，他们可以来和我进行私人的、没有任何记录的谈话，我可以帮助他们。所以，我所做的真的就是关心照顾他们，使他们对自己的工作保持兴趣，让他们看到团队的远景规划并参与其中。这真的花费我大量的时间：把大家组织到一起并让大家有一个共同的目标。因此，这3000个人并不是为我工作—可能20个人为我工作—而且我不会和所有的刚进团队一周的初级程序员一起讨论团队的规划。我和高级技术管理者—最上面两层的管理者讨论，然后他们再去管理他们的团队。

**玛丽安：**3000个人有你的家庭电话，你多久会接到一个电话？

**帕特：**也就是一周一到两封电子邮件的事。如果我没有去旅行，我几乎每一顿午餐都会和一个人开指导会议。我一般要直接指导30个人。而且我拥有一大批“有问题随时讨论”的员工。

**玛丽安：**那你直接指导的人是来自某些职位或者某些部门吧？

**帕特：**他们当中有一些是新人，有的进公司才两年。但是我认为他们很有天分。还有一些进公司的时间差不多和我一样长。

**玛丽安：**是什么促使你成为一个管理者而不是一个纯粹的研究者，你是如何转变的？你的意见会对有一些想转变的人提供帮助。

**帕特：**我在职业生涯早期就做出了这个决定，现在看来可能有点太早了。我在IBM工作了三年并且刚刚做完 System R，那时我刚成为一个项目经理，并且像你刚刚提到的一样，同Bruce, Jim和Franco一起工作。而我决定做管理的一个重要原因是我发现和一个团队的人一起工作得到的成就要比只靠我这一双手多得多。这就是我的动机。我用三年时间做一线产品经理，做R\*分布式数据库项目，然后一跃成为四线管理者并用三年管理计算机科学部门。我认为上升的太快，虽然我很喜欢我的工作。

这就像在大学里修了22门课，因为计算机科学系有22个不同的项目，他们涉及的范围很广泛，从理论到磁盘驱动到打印机和数据库。这很有趣，可以你不能深入到其中某一项技术中去，我觉得我就是错过了这些。在管理计算机科学部时，我创建了数据库技术中心。在管理数据技术中心时我一直想，“我喜欢这项工作”。因此我去找我的老板并跟他说，“看见那项工作了吗？我想离开这项工作去做那项，因为我知道那项工作的远景并有信心实现它”。所以我开始了管理和技术结合的工作，我真的很喜欢。

**玛丽安：**那么你对想从工程或者研究转向管理的人有什么忠告吗？

**帕特：**我今天想说的是，除非你在技术上取得了一些成就，否则不要做这种改变。一定要保证

你懂得什么是一个成功的项目和技术，因为这将是你以后工作的参考。因此，耐心的等一段时间，不要急于转向管理。善于积累你所做的工作和了解的技术，这将成为你的经验。然后你会有大量的时间进行转变。如果你善于管理和领导，那么每6个月就会有人想你提供管理的工作。

**玛丽安：**现在数据库研究领域有人认为IBM的数据库研究者有些过于以产品为导向。您是否注意到这样一种印象的产生？您同意这个看法吗？如果这是真的，那么如何保证IBM的研究不会错过改变数据库未来的机会？

**帕特：**在90年代中期，我们接触到Unix和Windows已经很晚了。我在多伦多找到Janet Perna，那时她正在负责开发Unix和Windows的数据库产品。我跟她说，“Janet，如果在Unix和Windows空间没有任何产品的话，那么我的研究团队的技术成果将不会为人所知。所以我想带着我们的技术帮助你开发数据库产品，并将其推向市场”。因此我们让研究者们参与产品开发，我想我们就是在那时改变了研究团队的本质。现在团队中仍然有些人很乐意从事产品开发和发布的工作。另外一方面，我们也让一些人回归到研究中，去做一些只有研究者做的事情。我们被认为以产品为导向的原因应该是我们的传统，正是这种导向使得IBM的研究团队保持竞争力和价值。

**玛丽安：**帕特，你的博士研究方向并不是数据库。从这一点来说，一个没有数据库研究背景的人可以转到数据库研究上来吗？

**帕特：**我认为可以。特别是在一些领域例如隐私，它不需要拥有五年的预写日志或者并发模式的报告的深刻理论知识。

**玛丽安：**关系数据库研究领域一直坚持的基本原则是查询优化器应该产生能够优化任何查询的查询计划。可是工业界的用户导向查询优化概念看起来暗示着用户必须参与到优化过程中。这是否意味着研究领域没有兑现承诺？是否有更多的研究工作要做？

**帕特：**我认为查询优化器是我们所说的自主计算或者自我管理的第一次尝试。查询优化器已经有25年的发展历史，伴随着基于代价优化模型和相关执行技术的丰富与发展。我们会继续查询优化的相关工作。研究越来越好的模型、优化方法和执行技术是我们永无止境的追求。因此模型越是能够预测数据的组织方式和处理流程，我们的系统就越接近于理想系统。

我们看看优化器提示的概念。现在提示的主要问题是人们无法直接输入一个即席查询，而是让数据库管理员进入机房并大喊“先别按‘输入’键，别写最后的分号，让我把提示加入到你的查询中再开始查询。”很多打包好的应用，像PeopleSoft, SAP, Seibel Systems，他们都是预先包装的。你无法得到这些查询，无法加入提示。查询语句越来越多，目前没有办法随时随地为用户定义的查询加提示，因此我们一定会坚持投资这项技术。

**玛丽安：**有没有一种方式使得提示能够传递给预编译的包，这样他们能够在运行过程中处理？

**帕特：**你总有让软件供应商加入提示的方法。我觉得他们并不愿意去实现它。作为一个独立的软件供应商，我并不想去学习某一种查询情况下最好的查询计划是什么，而且我不知道将来数据表的规模会变成多大。

**玛丽安：**我想象的是在运行过程中，一些统计信息和数据分布信息可以传送给软件包，然后这些包能够编译并产生正确的查询计划---

**帕特：**实际上就是这么做的。我们从引擎目录中得到数据的统计信息并用于基于代价的查询优化。因此，从某种程度上来说，如果模式运行正常的话，你已经得到了提示。

**玛丽安：**喔，我原以为软件包已经解析完查询并准备执行，我们无法再改变他的执行方式。

**帕特：**不，他们不是一种预先设定好的执行模式。他的执行模式是，“执行该字符串”。然后基于你的特定数据的统计信息的查询优化开始起作用。

**玛丽安：**我这儿有一张你早期在System R工作时的照片，你的性格看起来很阳光，这显然很不符合70年代计算机研究者的刻板形象，没有油腻的头发，没有计算尺，甚至没有眼镜。多年来，一个不符合刻板形象的人加入计算机行业变得容易了还是困难了？

**帕特：**我觉得——我希望我仍然看起来像那时一样——今天人们比以往更容易接受多样性。实际上我们也在通过不同的方式追求多样性，因为我们需要注入新鲜的想法。所以我很期待有一些数据库以外背景的人带来其他领域的经验。我觉得这样会带来生活方式、世界观以及文化背景的多样性。

**玛丽安：**你知道，我原来从来没有想过，不过既然你已经提到了他们最初让你管理Jim Gray和Bruce Lindsay——我看过他们那个时候的照片，我觉得这对一个这么干净、温柔的人似乎有些残忍。

**帕特：**这个年轻的女孩，是的。我们实际上很般配，我们都有长的金发。

**玛丽安：**好的。

**帕特：**所以我们没有多样性。

**玛丽安：**如果给你多余的时间去做一件在工作上没有做过的事情，你会做什么？

**帕特：**你知道，我一直希望能够更多的时间去阅读数据库领域以外的书籍。我一直没有足够的时间去做，但是我真的很想做。

**玛丽安：**这种情况会变得更糟糕，如果我们有越来越多的会议，越来越多的数据库研究者和有意思的材料要阅读。

**帕特：**是的，当然。

**玛丽安：**作为一个计算机科学研究者，如果你有机会改变自己，你将改变什么？

**帕特：**我想我会了解更多关于硬件的知识。

**玛丽安：**硬件？

**帕特：**是的。我想这是我很想学习的一个领域。

**玛丽安：**它会如何影响你所做的事情和你的想法？

**帕特：**我觉得现在数据库面临着一个现实，就是处理器已经快了两倍而磁盘仅仅快了5%到6%。因此突然间性能出现了很大的差距。我希望我们有方法来填补这种差距，例如通过避免重复I/O和缓冲区预取。我们有很多技术并且正在发明更多的技术来减少性能差距。我想如果我们懂更多硬件的知识并且和处理器I/O、存储系统的人多交流，我们可能会有更多利于每个人的发明。

**玛丽安：**你知道，我想对于一个刚毕业的数据库博士，他懂得磁盘工作细节的可能性几乎为零。

**帕特：**是的。

**玛丽安：**John Wilkes有一篇描述磁盘细节的论文，在这个领域发展了十年或者十五年以后我再读它，有很多地方看不懂。也许受过数据库教育的人在理解那个磁盘盖子下面的东西时候会有一些困难。

**帕特：**是的，随着这个领域的发展，越来越多的知识被抽象化。今天我们头脑中有一种简单的模型，一个磁盘就是一个传动手臂和一个盘片，它们存储数据库的数据。而实际上存储数据的是磁盘阵列，是存储区域网络，是在那个磁盘盖下面的各种不同的结构。它们被一个逻辑卷管理器包装起来，而这些逻辑正是由操作系统领域的人设计，它们可能懂也可能不懂数据库。有时候这种透明性是很好的，让我们更专注于产品和细节。现在我正在开发自我管理数据库，所以我很高兴他们在开发自我管理文件系统。从另一方面来说，优化整个栈会更好。所以两个领域的人应该多交流，我们也希望尽快地使用他们的新技术。

**玛丽安：**非常感谢你，帕特。

**帕特：**别客气，这是我的荣幸。

(史英杰译，马如霞校)