CRISP-DM was conceived in late 1996 by three "veterans" of the young and immature data mining market. DaimlerChrysler (then Daimler-Benz) was already experienced, ahead of most industrial and commercial organizations, in applying data mining in its business operations. SPSS (then ISL) had been providing services based on data mining since 1990, and had launched the first commercial data mining workbench – Clementine – in 1994. NCR, as part of its aim to deliver added value to its Teradata data warehouse customers, had established teams of data mining consultants and technology specialists to service its clients' requirements.

At that time, early market interest in data mining was showing signs of exploding into widespread uptake. This was both exciting and terrifying. All of us had developed our approaches to data mining as we went along. Were we doing it right? Was every new adopter of data mining going to have to learn, as we had initially, by trial and error? And from a supplier's perspective, how could we demonstrate to prospective customers that data mining was sufficiently mature to be adopted as a key part of their business processes?

A standard process model, we reasoned, non-proprietary and freely available, would address these issues for us and for all practitioners.

A year later we had formed a consortium, invented an acronym (CRoss-Industry Standard Process for Data Mining), obtained funding from the European Commission, and begun to set out our initial ideas. As CRISP-DM was intended to be industry-, tool- and application-neutral, we knew we had to get input from as wide a range as possible of practitioners and others (such as data warehouse vendors and management consultancies) with a vested interest in data mining. We did this by creating the CRISP-DM Special Interest Group ("The SIG", as it became known). We launched the SIG by broadcasting an invitation to interested parties to join us in Amsterdam for a day-long workshop: we would share our ideas, invite them to present theirs, and openly discuss how to take CRISP-DM forward.

On the day of the workshop, there was a feeling of trepidation among the consortium members. Would no-one be interested enough to show up? Or if they did, would they tell us they really didn't see a compelling need for a standard process? Or that our ideas were so far out of step with others' that any idea of standardization was an impractical fantasy?

The workshop surpassed all our expectations. Three things stood out:

- Twice as many people turned up as we had initially planned for.
- There was an overwhelming consensus that the industry needed a standard process, and needed it now.
- As each attendee presented their views on data mining from their project experience, it became clear that although there were superficial differences – mainly in demarcation of phases and in terminology – there was tremendous common ground in how they viewed the process of data mining.

By the end of the workshop, we felt confident that we could deliver, with the SIG's input and critique, a standard process model to service the data mining community.

Over the next two and a half years, we worked to develop and refine CRISP-DM. We ran trials in live, large-scale data mining projects at Mercedes-Benz and at our insurance

sector partner, OHRA. We worked on the integration of CRISP-DM with commercial data mining tools. The SIG proved invaluable, growing to over 200 members and holding workshops in London, New York and Brussels.

By the end of the EC-funded part of the project – mid-1999 – we had produced what we considered a good-quality draft of the process model. Those familiar with that draft will find that a year on, although now much more complete and better presented, CRISP-DM 1.0 is by no means radically different. We were acutely aware that, during the project with the process model still very much a work-in-progress, CRISP-DM had only been validated on a narrow set of projects. Over the past year, DaimlerChrysler has had the opportunity to apply CRISP-DM to a wider range of applications. SPSS' and NCR's Professional Services groups have adopted CRISP-DM, and used it successfully on numerous customer engagements covering many industries and business problems. Throughout this time, we have seen service suppliers from outside the consortium adopt CRISP-DM; repeated references to it by analysts as the de facto standard for the industry; and a growing awareness of its importance among customers (CRISP-DM is now frequently referenced in invitations to tender and RFP documents). We believe our initiative has been thoroughly vindicated, and while future extensions and improvements are both desirable and inevitable, we consider CRISP-DM Version 1.0 sufficiently validated to be published and distributed.

CRISP-DM has not been built in a theoretical, academic manner working from technical principles, nor was it created behind closed doors by elite committees of gurus. Both these approaches to developing methodologies have been tried in the past, but have seldom led to practical, successful and widely–adopted standards. CRISP-DM succeeds because it is soundly based on the practical, real-world experience of how people do data mining projects.

Although CRISP-DM is excellent for what it is intended for, it cannot solve all challenges facing the data mining industry. CRISP-DM facilitates planning, documentation, and communication in data mining projects, and it can serve as a common reference framework for the data mining industry. But it is, by design, a "piece of paper". It can help the communication between people but it lacks the formal rigor necessary for communication between software systems. Other efforts represented on this panel address some of the communication issues between (data mining) systems, but none of them alone is a solution for all relevant problems either.

Therefore, the CRISP-DM consortium welcomes the KDD 2000 panel as an opportunity to discuss the contributions of various standardization efforts, and hopefully, this panel will mark the start of a joint effort. CRISP-DM can contribute to such a joint effort by delivering the common framework where all relevant interface standardization efforts, already existing or not, can fit in.

The CRISP-DM Consortium
August, 2000