Spatial data for the oil industry:

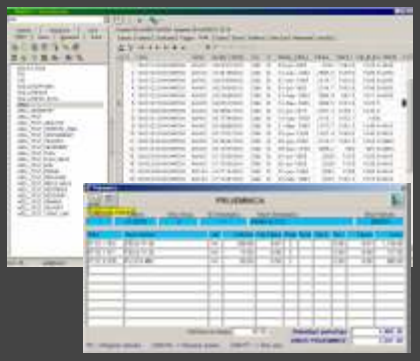
How to do it and where to get it

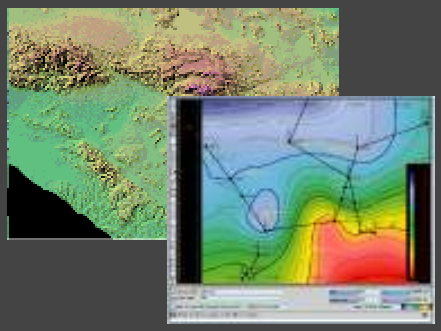
Ian Batty
Maguffyn Consultants

ESRI Canada PUG, Calgary 11-Apr-2003

Bridging the divide

- Existing relational data management
- Spatially integrated data management





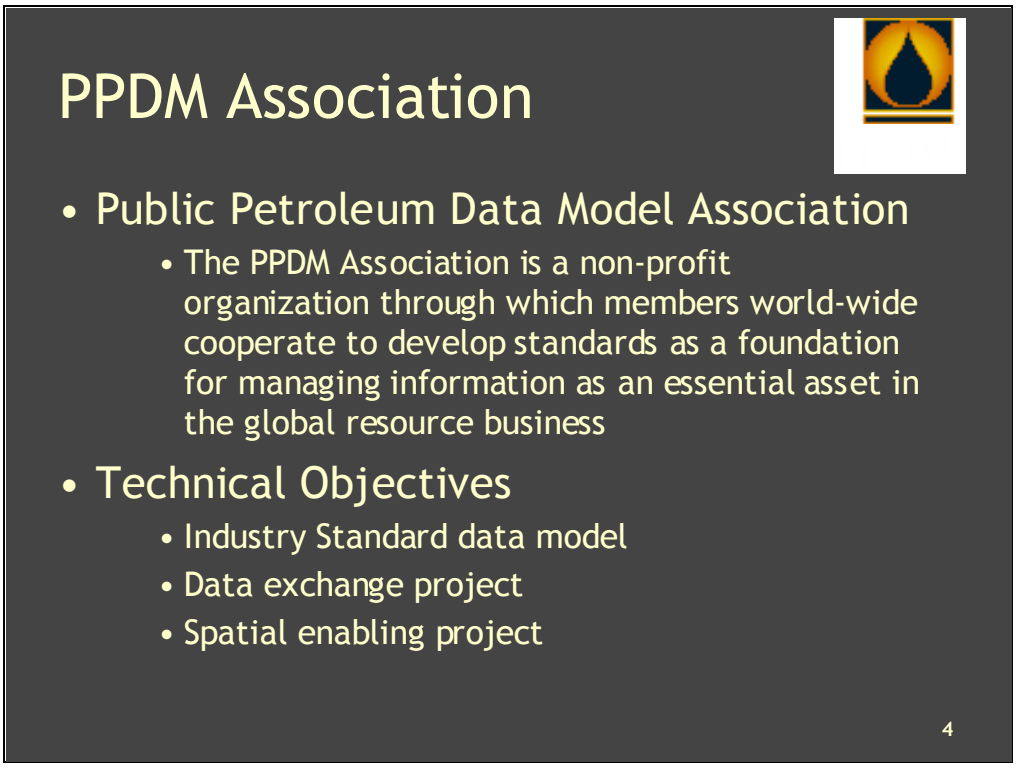
2

What data is spatial?

- Just about everything in the oil industry has a spatial component
- Even legal documents like contracts often cover exploration or development for an area
- GIS systems have not been used to find corporate (enterprise) data- Because it is hard to link relational data and a GIS

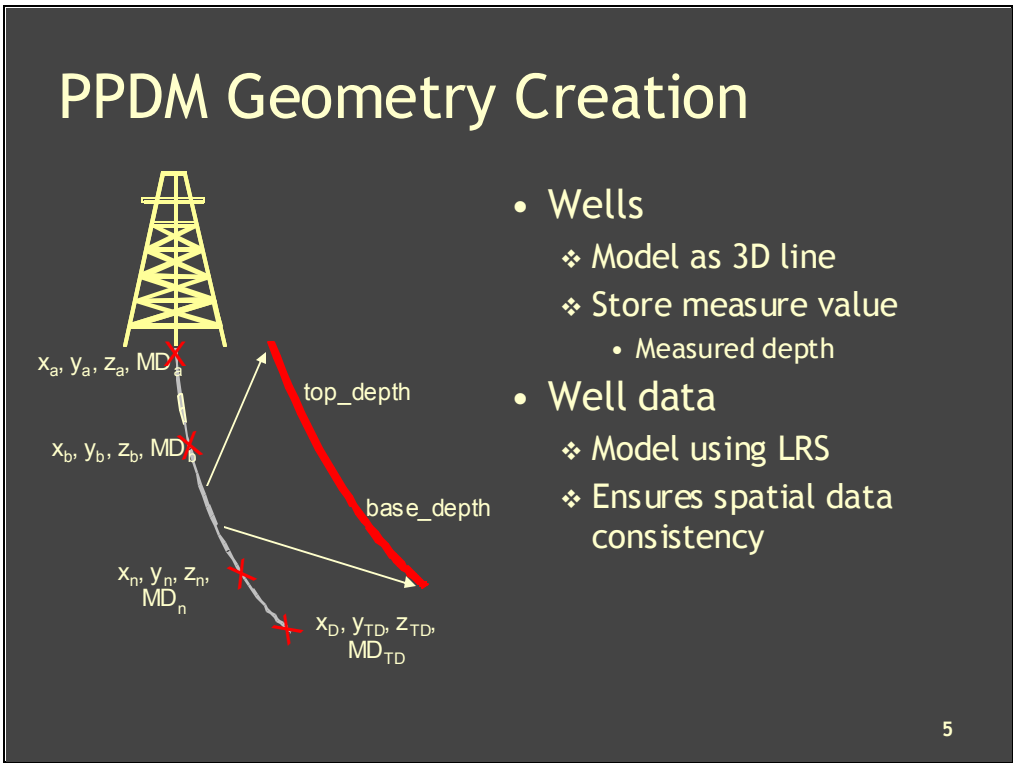
PPDM & ESRI Canada

- We need a little background to understand how PPDM fits into the GIS world.
- ESRI has developed a number of data models for different vertical industries that allow people to move from file based data systems into the Geodatabase. The challenge facing the Petroleum Industry was how to move from an existing data model to a Geodatabase.
- Back in 2001 people began to think that creating this standard Geodatabase model would be a good idea, but being the real world it took another PUG and a lot of background work to actually get cheques signed so that resources could be obtained and start the technical work.
- ESRI Canada provided technical resources and software to the project as well as financial support.
- By October of last year the first set of deliverables were made available at www.ppdm.org/products/spatial



Background

- The PPDM is the Public Petroleum Data Model Association- it has a mission statement and a number of technical objectives. The PPDM is a physical implementation of a data model that represents data and data relationships used in E&P. The objective that is particularly pertinent to this presentation is the Spatial Enabling Project.
- Some of you may be some of you intimately acquainted with PPDM, some only vaguely familiar, and a good group of you are probably wondering "what does a data model organization have anything to do with the PUG?" or "Didn't these guys go out of business years ago?". The answer to the second question, quite obviously, is no. The answer to the first question is a little more involved but will hopefully be answered by the time I stop talking.
- The key thing to note about the Projects, as opposed to the data model, is that they are completely OPEN to all and ANYONE to participate in. If you want to set the direction that the project takes, step up to the plate and sponsor the project. If you want to observe, come and attend the meetings. If you can't attend, sign up to the e-mail list- you will receive the presentation and notes from the meeting. But if you just want to use the results of the project, then that is cool too- but to gain the real benefit you may need to join the Association and be a member of the PUG.
- Oh, and the final thing to note about the Projects is that as well as producing guidelines, recommendations and methodologies on how to do things, they also produce real live deliverables that are actually implemented in a technology (or 2). The methodologies are as open as they can be- the implementation, by its very nature has to pick a technology.



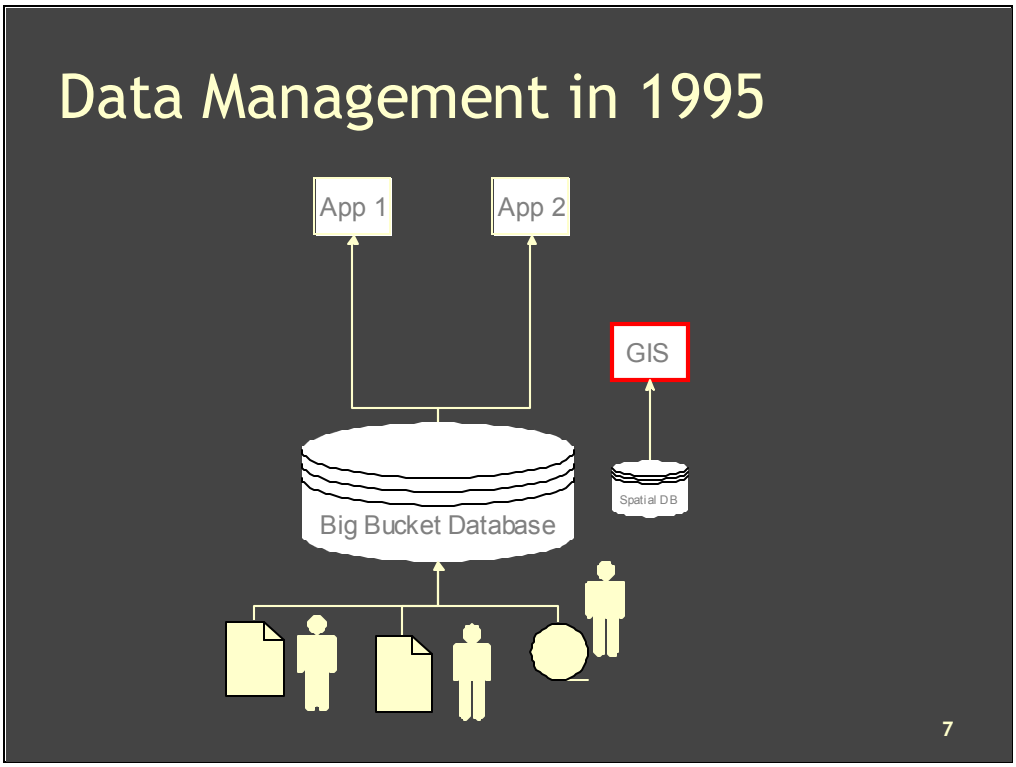
Spatial Enabling Methodology

- The first problem that people had was simply the problem of how do you get data from a relational database and into a GIS?
- As wells are quite important to the oil industry we tackled them first and have created a methodology that allows well data in a relational database to be modeled in UML and then the data transformed via some software into a spatial database.
- The basic approach is to model a well as a 3D line with measures. The measure value is the measured depth. This allows all data in the well module that has a MD value associated with it to be displayed on a GIS. The business tables themselves do not need to be spatial enabled, the geometries do not need to be maintained, but the data can still be displayed.
- However, just telling people how to do something is not as useful as actually doing it. So you can also download the scripts to create the geometry and can even get sample databases that have been spatially enabled.
- If you don't want to use a full Geodatabase but need cross application capability too- then that is catered for as well by creating an OGC compliant simple feature database. But the real live implementations of these 2 options are in SDE Binary and Oracle Spatial storage formats.
- To allow people who are not familiar with a PPDM database we also provided some high level UML diagrams for the well header and well test module. And finally, we provided a detailed UML document that ArcCatalog can read to create a Geodatabase.

Geometry Implementation

- Java scripts
  - Available to PPDM members
- Other software solutions
  - FME
- Storage format neutral
  - SDE Binary
  - Oracle Spatial

6

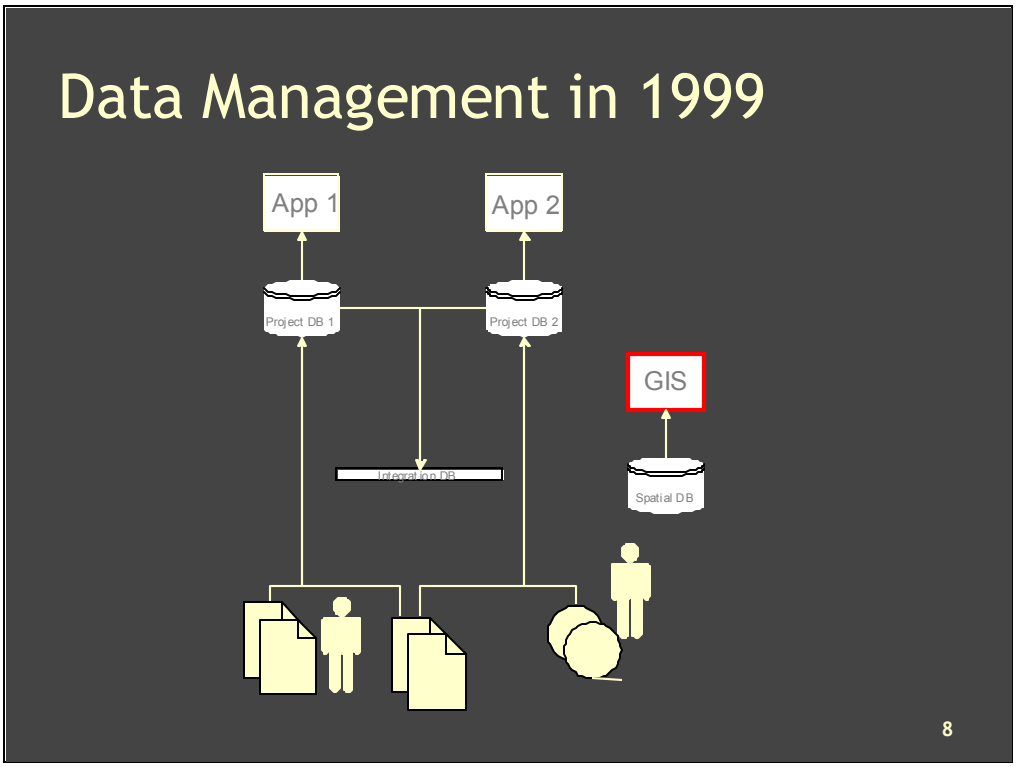


Introduction

- But enough of the stuff that we don't know about- the only way to get spatial data accepted is to show that it is very important and that it needs to be incorporated into your data management practices. To understand why I believe the data isn't incorporated I want to give brief overview of the evolution of data management, purely as I see it, over the last few years.

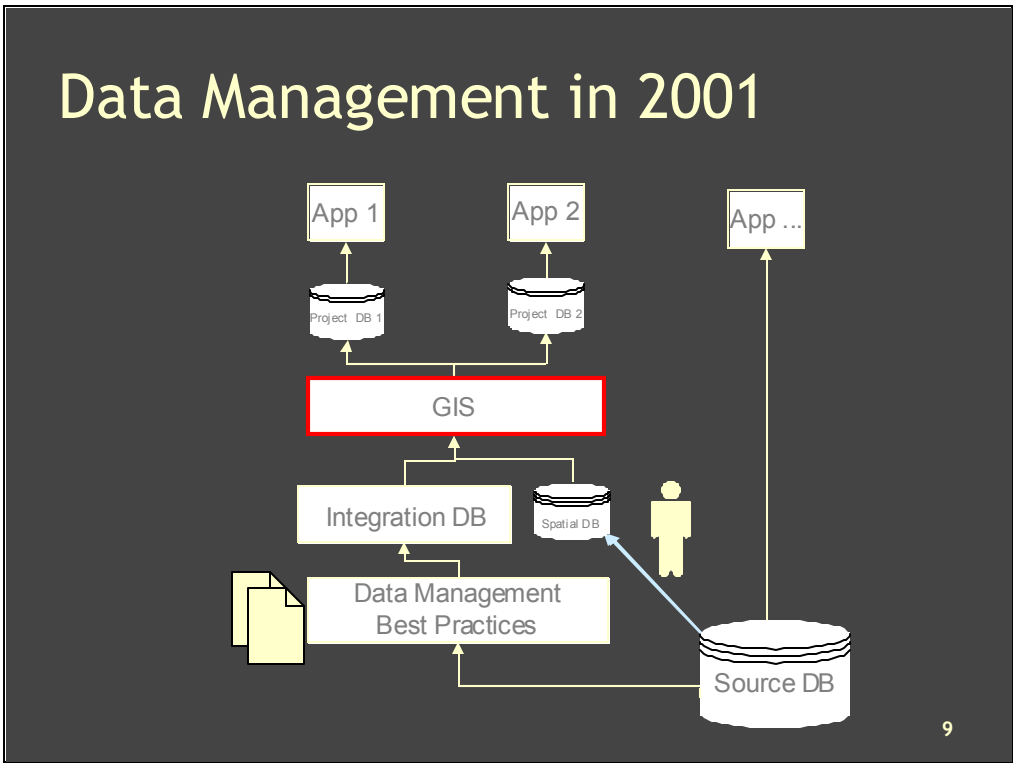
Slide body

- Back in 1995 we had lots of data, coming in in various different formats and a whole load of people to process it. Their main job was to provide the data to a series of applications so that users could use it. And wouldn't it be nice if we only had to deal with 2 applications. Obviously there were a lot more applications than 2 and so the thinking at the time was "let's put it all into 1 big bucket and then all the applications can work of that database" There were a couple of different options for the big bucket but that was the way to do it.
- Guess what? The big bucket didn't work. The experience of a large number of people has shown that if you are still trying to use a big bucket today you are in for a really hard time.
- Spatial data? Oh yes, spatial data was doing its own thing off to one side.



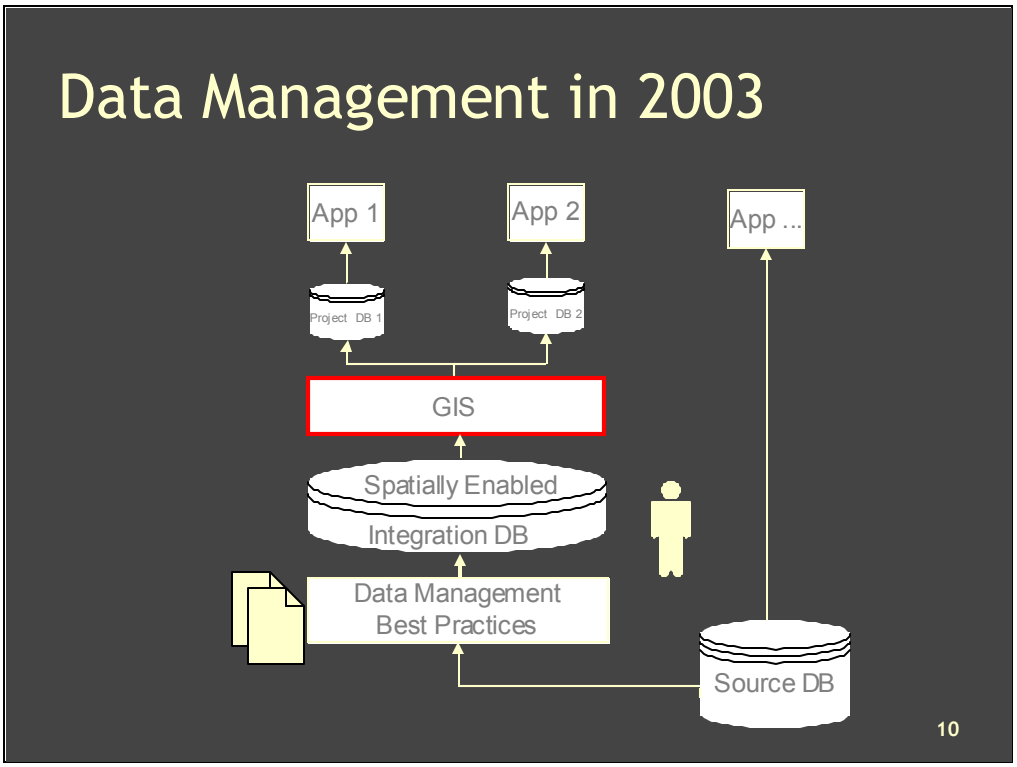
Slide Body

- So we move onto 1999. We still have data arriving in lots of different formats- in fact we have more data arriving than ever before. But in a bout a true logical thinking- there has been a round of layoffs and we don't have as many people working on that data.
- We still have the applications at the top of the picture- but the various vendors, bitten from the failure of the big bucket, have gone away and created their own project stores. Some have even created integrated project stores. But even the integration doesn't address all the functionality, so we still have multiple stores.
- And the end-users are happy.The data management department is working overtime having to load all this data into multiple data sources. And there is a little itch in the back of the minds of the data management people saying "we have to make these stores talk to each other"
- The idea that people came up with was to make a little integration layer- but it should be as thin and small as possible- just store the cross references and the system will work. The databases will talk to each other and we won't duplicate any data.
- Except that when we came to build these systems they got so complex that they became basically unmaintainable. I know, I tried to build one.
- Meanwhile, spatial data was doing its own thing, off to one side.



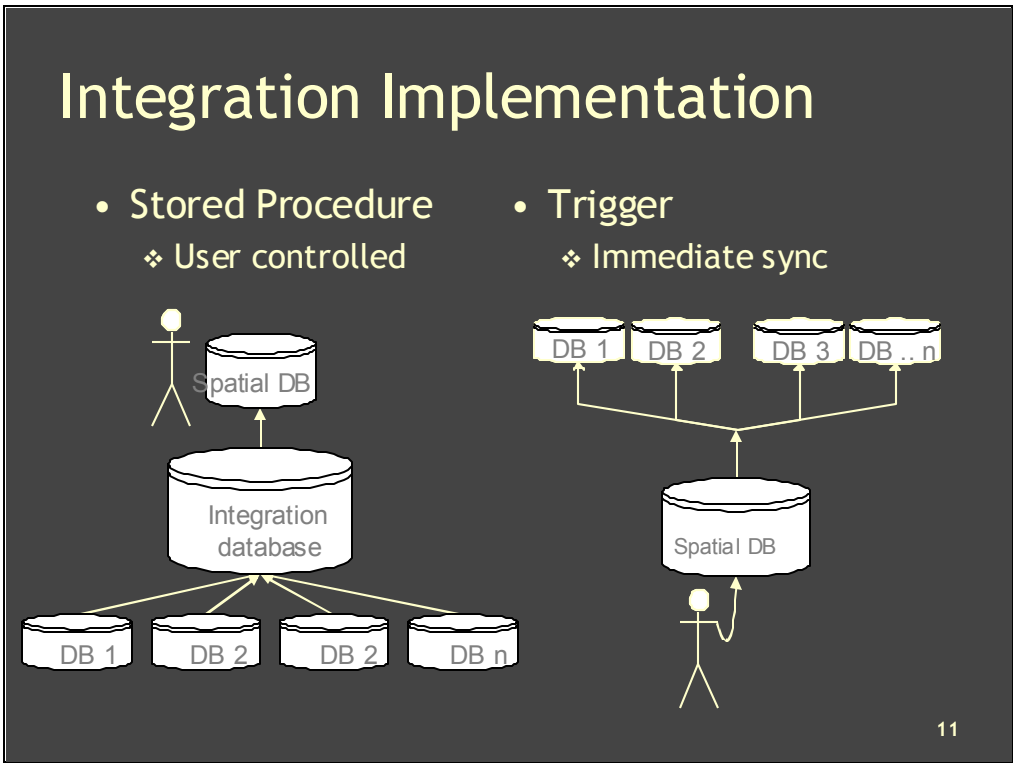
Slide Body

- So we move on to 2001. The world has changed- we are now getting our data from a vendor who does all the formatting that the data management department used to do. So now we only have 1 person working there- but he knows a LOT about the database.
- The applications are still being used by the end user and they are all working from their own project stores. The problem facing the users now is physically finding the data and knowing when new data arrives or is updated.
- GIS is a GREAT way of doing that- people can look at a map and in seconds know if there are any changes or if things need to be updated. So we need to move the GIS system into the main data flow. And it can be used to decide which data is moved into the project stores.
- So that needs a spatial database in the main flow. And fortunately the vendor can supply the data so that we can load it straight into the project store.
- But you know that the world is never straightforward. The data management department, in their infinite wisdom have created a series of "Best Practices" and all the data has to go through these best practices. The problem is that the data management department understands attribute data really well- so that can through the best practices, but they don't understand spatial data. So we finish up with the source data going straight into the spatial database.
- The other issue that we face is that because the source data is coming from a number of different sources, all with best practices different from our own we need to store the attribute data in some sort of integration layer. This layer has to store all the data that most users will ever want to see and everything that can be displayed on a map.
- The alert people may be thinking "this looks a lot like a big bucket" The difference is that there are still applications that run of their own source data. Completely outside of the main flow. This was the failing of the big bucket- it tried to cater for every application- the integration layer is much more limited in its' ambition- it just wants to cater for most of the people most of the time- but it also wants to allow you to know what data the other applications are using, even if it can't store the details about that data.
- So we finish up with a main flow and others. But the spatial data and the attribute data in the integration layer don't really talk to each other. That itch that the data management department had? It is still there. Not as big, but itching none the less and the fear of data busts is just as large as ever



Slide Body

- Which brings us up to date. The good news is that unlike the real world, the data management world has not changed that much since 2001. Source data still arrives, applications exist, they use their own data stores, people still like to use a GIS to understand the data and best practices still cause as much trouble as ever, and I should add, still solve as many problems as ever.
- The only difference is that there is a methodology to integrate your spatial data with your attribute data and create a single, consistent, clean, contained integration layer.



Stored Procedure

- The integration database contains all the information to map different entities in different databases and ensure consistency.
- Stored procedures are generally used when the end user requires control over the timing of synchronization and does not have control over the quality of the data that is available to him. It is also used where the volume of data being transferred is large or when the synchronization occurs at regular intervals and can be scheduled to occur out of office hours.
- The advantage of using stored procedures is that it can run once and can synchronise many databases. It also uses code that is centralized.
- The disadvantage is the currency of the data depends on the frequency of running the stored procedure.

Trigger

- The user uses the data management processes defined by the data management group. This means that the source database uses the correct naming conventions etc so that there is no need to use a metadata model to map the names
- Triggers can be used when the data is of known (i.e good) quality, or when the volume of data is small.
- The advantage of using triggers is that currency of the data is guaranteed: as soon as the change is made in the source database it is replicated to all other databases.
- The disadvantage of using triggers is that the user may feel that they have lost control over their data. In addition the triggers have to be implemented on each database that is a source.

PPDM Technical Deliverables

- Software and scripts
  - Geometry generation for wells
  - Data integration
- Sample dataset
- UML for Geodatabase creation

12

PPDM Objectives for 2003

- Capturing object behaviour
  - e. g. Assign well to field automatically
- Modeling objects through time
  - e.g. Licensing rounds
- Modeling objects with depth
  - e.g. Horizons
  - e.g.Pools

13

IPL Spatial Database

- Uses PPDM methodology
  - Geometry generation
  - Data integration
- First production implementation
  - Full western Canada dataset

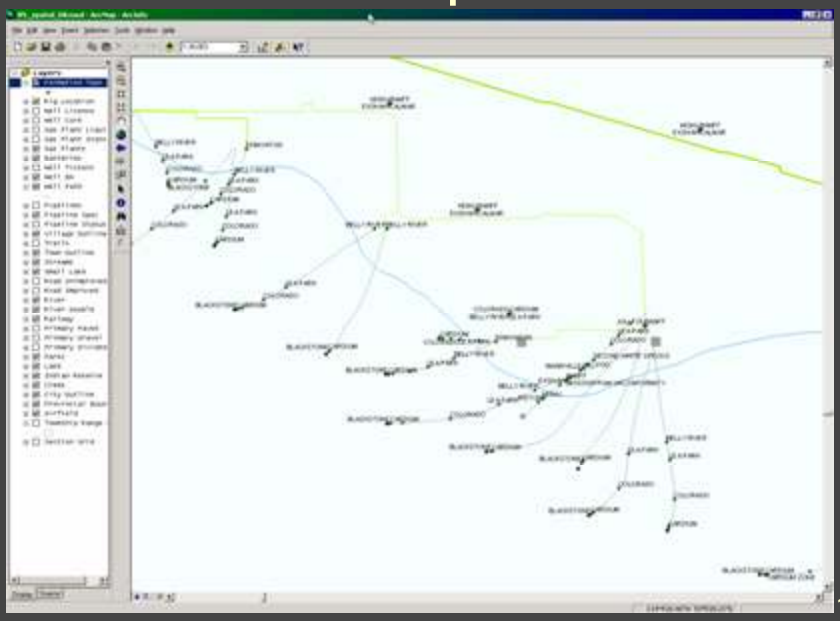
14

International Petrodata Ltd

- Data vendor
  - Established in 1963
- Data source
  - Petrodata File
- Digital data availability
  - Petrodata network

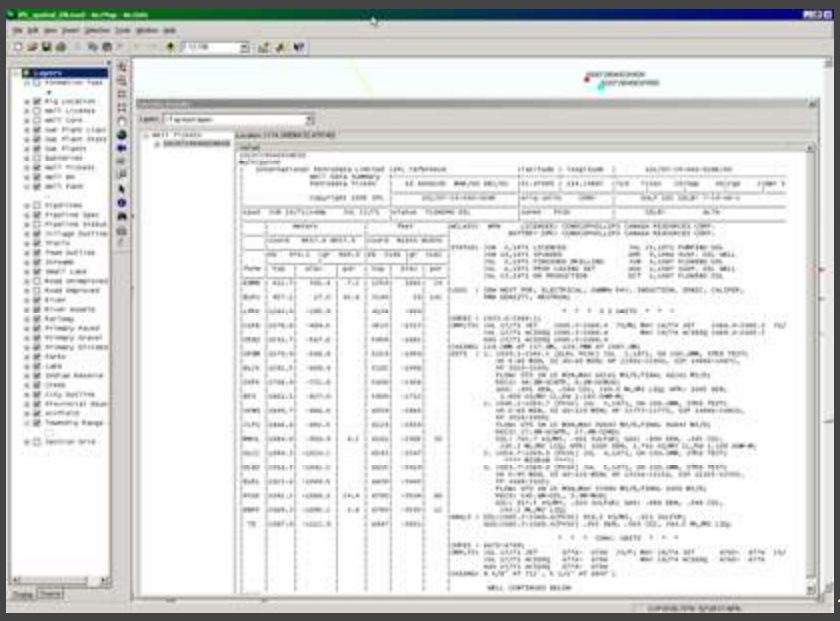
15

Well formation tops



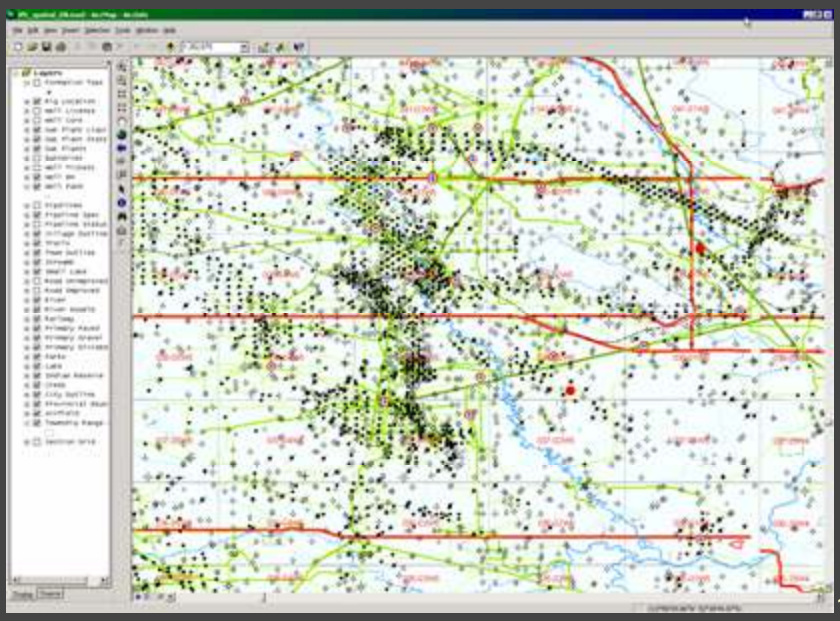
16

Well Ticket



17

Data available



18

Performance considerations

- Oracle
  - 3.5GB RAM allocated to Oracle
    - Entire database in RAM!
- Storage format
  - SDE Binary
- Spatial view optimisation
  - Functions to extract relevant data
  - /* ORDER */

19

Contact Details

- e-mail
  - ian@maguffyn.com
- PPDM
  - curtist@ppdm.org
    - www.ppdm.org/products/spatial
    - www.ppdm.org/development/projects/spatial
- IPL
  - brad@petrodat.com
    - www.petrodata.com

20