

Script for Oil and Gas Governance Demo

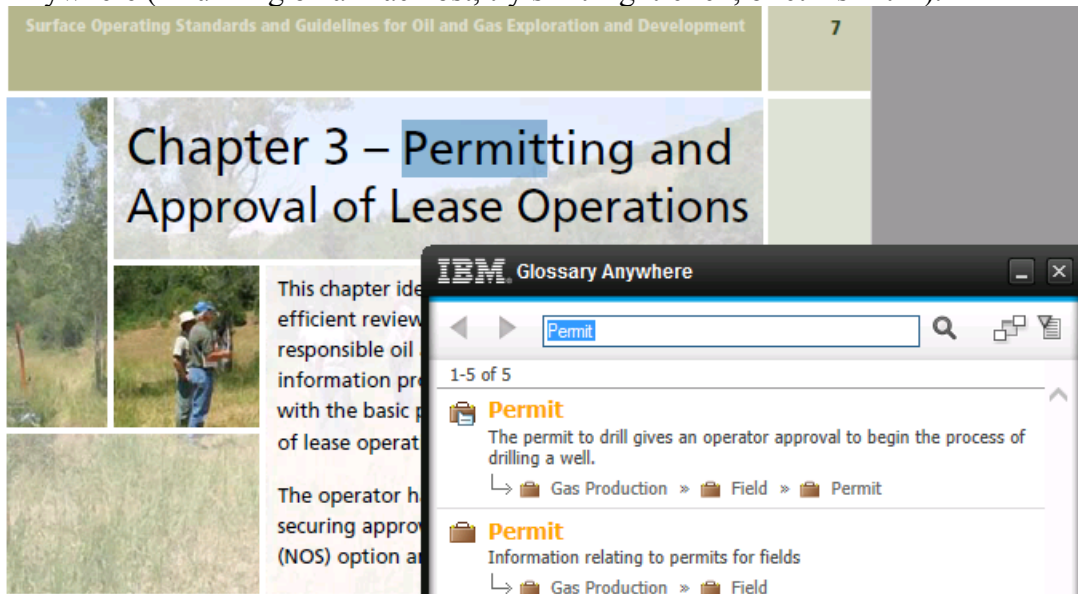
Send feedback to: neil.williams@us.ibm.com

Setup

1. This demo typically runs 45 minutes but can easily go 90 minutes with questions and diversions.
2. Prior to start of demo perform these steps to speed things up:
 - a. Open “BLM Gold Book” PDF on desktop, jump to chapter 3.
 - b. Highlight “Permit” in chapter heading and hit ctrl-F10 (ctrl-shift-L on Mac) to bring it up in BG Anywhere, then close BG Anywhere (you’ll open it during the demo but now the results will be cached and come up faster).
 - c. Open a browser and open IIS Launchpad, launch IGC
 - d. Open a tab in your browser and go to Bluemix Data Connect. If working with CAP, I’ve found that using a browser directly on your host laptop vs inside the CAP image is faster when connecting to Bluemix.
 - e. Open a tab in your browser with Watson Analytics and go to your personal folder.

Part 1: IGC

3. Open “BLM Gold Book” PDF on desktop, jump to chapter 3. Set the scene with you as a business analyst or data scientist looking at any typical document – PDF, Excel, Word, Cognos report, etc.
4. You’re interested in what types of data and rules your company has around Well Permits. Highlight “Permit” in the chapter heading and hit ctrl-F10 to bring up BG Anywhere (if running on a Mac host, try shift-right-click, or ctrl-shift-L).



5. Describe BG Anywhere use and audience. BG Anywhere is a lightweight applet that can be given to anyone in the company that needs simple lookup access to the glossary.
6. Click on the term: Permit (the term, not the category/folder).
7. Briefly describe the information shown.
8. Explain that you're a business analyst or data scientist rather than a casual user and want to explore the information in more details.
9. Click on link to open in browser to pull up the full IGC interface.
10. Describe the attributes in more detail. Point out link to the PDF doc. This is a custom attribute and customers can put their links and docs in these fields. If you put this term in a collection, point that out and explain that collections can be used for groups with a shared interest or in a specific line of business.
11. Point out the Cognos report (fyi this is a Cognos report from the JKLW financial demo), explain that report metadata can be governed in IGC.
12. Hover over the Cognos report and bring up the technical lineage. Explain that users have their favorite reports, but frequently need to modify the report and data to suit their needs. With data lineage, they can see how the report is created and what elements they can use to create their own datasets.
13. Explain that the lineage for this report is complex, but there is a business lineage report which can be simplified/customized to only show elements of interest to various classes of business users. Exit the lineage report.
14. As a data analyst, we're interested in how the information in the glossary is structured and what's available at a high level. Go to the category hierarchy (under the Glossary menu) and explain what categories are, and show how the Gas Production category is organized (i.e. Fields have Permits and Wells).
15. Go back to the Permit term.
16. Point out the Permit Number Exists rule.
17. Click on the rule and explain the governance policy hierarchy and how companies can use it. For example, the legal department might create business policies around Government Obligations, with a specific group of Government Regulations the company must follow. Those regulations flow to specific governance rules such as making sure fields have permits. IGC provides the linkage of these business policies to the technical assets that ensure rules are followed. In this case, there are two data rules, PermitNumberExistsPermitTable and PermitNumberExistsWellTable which check the database to ensure permit numbers exist for fields and wells. A graph of the logical hierarchy is at the bottom of the permit term screen.
18. Back up to the Permit term. Hover over the assigned technical asset `GAS_PRODUCTION_DATA`. Point out that this is a database table with information on oil fields and as a data scientist this is the information we're interested in. However, we want to create our own copy with a modified set of data so we're interested in how the `GAS_PRODUCTION_DATA` is populated.
19. Click on the `GAS_PRODUCTION_DATA` table and point out the link to Bluemix Data Connect for performing end-user ETL. Point out the columns in the table, hover over a column to show the associated term and definition – explain that this can be used by end-users to understand what type of data is in particular columns. Point out the quality analysis information, including the number of rows and data quality score

of 99%. Since the quality score is high, we can trust the data in the table. [If you don't have an analysis link, run and publish an analysis for GAS_PRODUCTION_DATA in the IA thin client]

- a. Optionally show the IA thin client if the customer is interested in the data quality scoring and data profiling.
20. Bring up the technical lineage for GAS_PRODUCTION_DATA.
21. Point out that GAS_PRODUCTION_DATA draws from WELL_DATA and PERMIT_DATA, and they flow through an ETL job called LOAD_OIL_DW.
22. Click on the "expand" link in the ETL job. Point out that the ETL job is doing a simple join on the PERMIT_NUMBER column.
23. We're interested to see if PERMIT_NUMBER is a unique value or primary key, so let's look at the PERMIT_DATA table. Close the lineage windows.
24. Click on the PERMIT_DATA table from the Permit term screen.
25. We see that there is a primary key for the table. Hover over the primary key and we can see that it applies to the PERMIT_NUMBER column, so we can use that for our join.
26. We're also interested in exploring what information we have for Wells. Back up to the Permit term, then click on the Well term (under Associated Terms).
27. Click on the Fracking Fluid class and show the Valid Values. Explain that custom classes can be created for enforcement or to measure data quality.
28. Now you're satisfied that you understand the data and which tables you need, so you can switch to Bluemix Data Connect. Users can click the Data Connect link previously shown, but explain that you've already logged on in a separate window to save time.

Part 2: Bluemix Data Connect

1. Show the types of connections supported
2. Show existing activities, explain that this is how Data Connect ETL jobs are organized and that groups of users can share the same space and see/edit the same activities.
3. Show Refine & Copy interface
 - a. Connect to your data source and select the PERMIT_DATA and WELL_DATA tables
 - b. Show DQ score for table and columns, point out the automatic suggested changes to improve quality. The WELL_DATA table should have several suggestions.
 - c. Show filtering, standardization, etc
 - d. Show how to join the two tables, point out that thanks to IGC we know to use the PERMIT_NUMBER column
 - e. Show send to Watson target, but cancel without sending
4. Optionally show Design Data Flow interface for a more technical audience

Part 3: Watson Analytics

1. Switch to Watson Analytics and explain that you sent your dataset there previously. Explain that you also edited the column names in Data Connect to make them user-friendly. Thanks to IGC you were able to figure out what the typically cryptic column names were since IGC maps them to business terms with definitions.
2. Key in a query like “first 12 months gas by frac fluid type” and choose the “Packed Bubble” chart. Explain that Watson allows simple English queries and the cognitive capabilities interpret questions against the data to come up with suggestions.
3. Explain that this chart shows the effectiveness of different fracking fluids on the first 12 months of a well’s output. Point out the various drill-down options across the top of the screen.

Part 4: Conclusion

4. Conclude the demo:
 - You’ve shown how a business analyst or data scientist can start from a report, PDF, spreadsheet, etc.
 - From there they can research what the terms mean and which tables and columns to use for their own custom datasets, along with the quality scores for those tables.
 - Without having to call IT, line of business users can use Bluemix Data Connect to create their own custom ETL jobs
 - Data from their ETL jobs can be output to a Bluemix SaaS database like DB2 on Cloud, or sent to Watson Analytics (or many other targets including flat files).
 - Watson provides a quick, intuitive and simple interface for both regular analysis and predictive of the custom dataset.
 - All of these tools are available from a single vendor, IBM. All products are available on Cloud as a service