i2k Connect

Powering Knowledge Discovery and Knowledge Flow with AI

John S. Boden, SPE International
Reid G. Smith, i2k Connect Inc.

APQC's 2019 KM Conference

The vision of the Society of Petroleum Engineers is to enable the global oil and gas E&P industry to share technical knowledge needed to meet the world's energy needs in a safe and environmentally responsible manner. SPE is the industry's leading publisher of technical papers, best practices and lessons learned. However, the increasing volume of information and diversity of channels make it difficult to connect with the knowledge and subject matter experts needed to solve problems. As a result, cross-industry knowledge flow is impaired. To address the challenge, SPE has implemented a new research portal, supported by artificial intelligence (AI). The portal integrates subject matter expert knowledge with AI natural language processing and machine learning. It automatically enriches documents by classifying them into relevant taxonomies, geotagging oil fields, and extracting key concepts, authors, and institutions. These enrichments enable SPE members to zero in on the relevant information from all SPE channels and to graphically analyze timeframes, geography, related concepts and cross industry collaboration (using social network analysis).

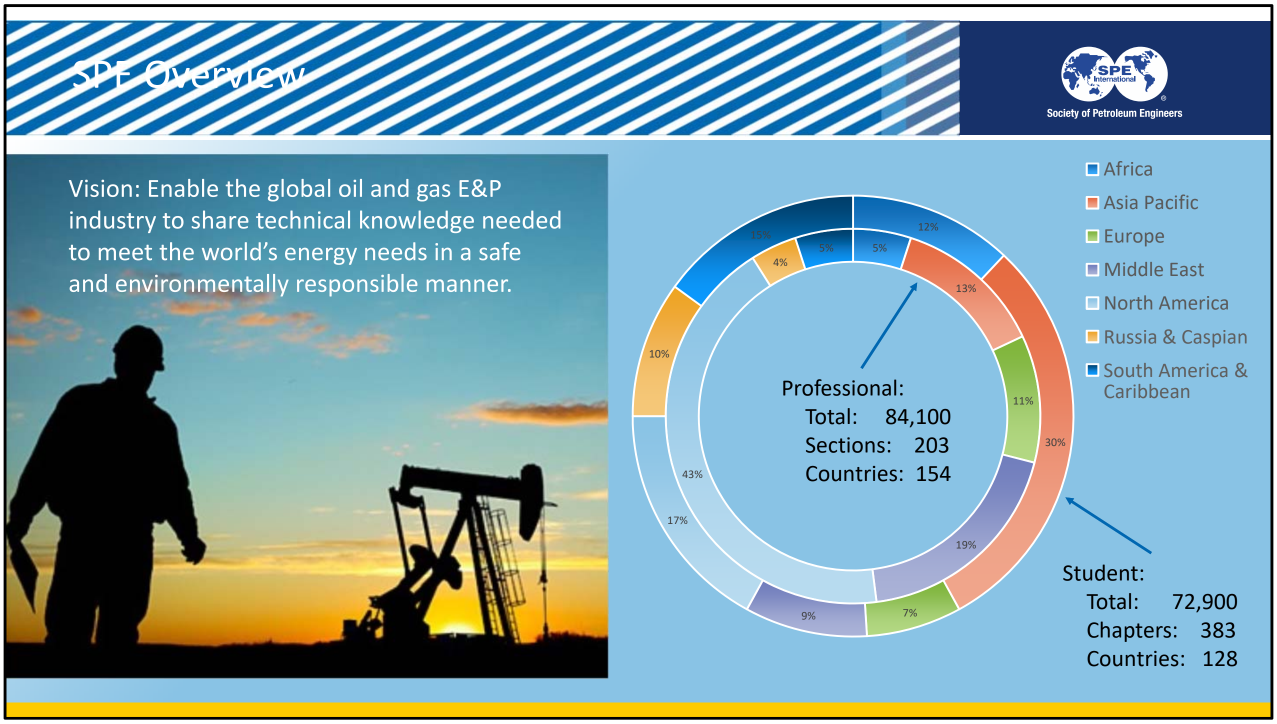
Keywords: AI, Cognitive, Tacit, Collaboration, Learning, SME, Social, Knowledge Flow, Classification.

Conference Theme: ***Bridging High-Touch and High-Tech.*** Explore the new capabilities that are affecting the KM discipline while positioning them within a holistic strategy that combines people, process, and technology. Learn which tools hold the greatest promise, how to build new KM tools and approaches into the flow of people's work, the keys to engaging end users, and how successful KM programs are weathering the storm of change.

Track: **High-Tech Solutions to Knowledge Problems**
Advice for navigating the evolving technology landscape and real-world examples of new capabilities that are improving knowledge discovery.

Outline of the Presentation

- SPE overview and the issues we set out to address
- SPE Research Portal
- Timeline/Implementation Process
- Results – Measures of Success
- Lessons Learned



E&P = Exploration & Production

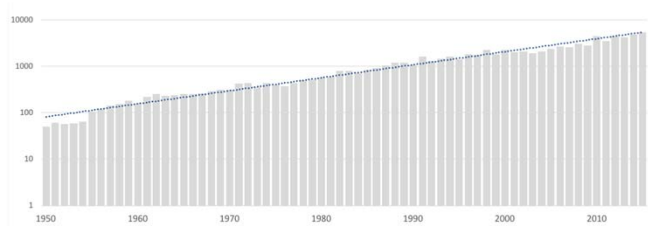
Mission: collect, disseminate, and exchange technical knowledge concerning the exploration, development and production of oil and gas resources, & related technologies for the public benefit; and to provide opportunities for professionals to enhance their technical and professional competence.

Vision: enable the global oil and gas E&P industry to share technical knowledge needed to meet the world's energy needs in a safe and environmentally responsible manner.

Addressing the Issues

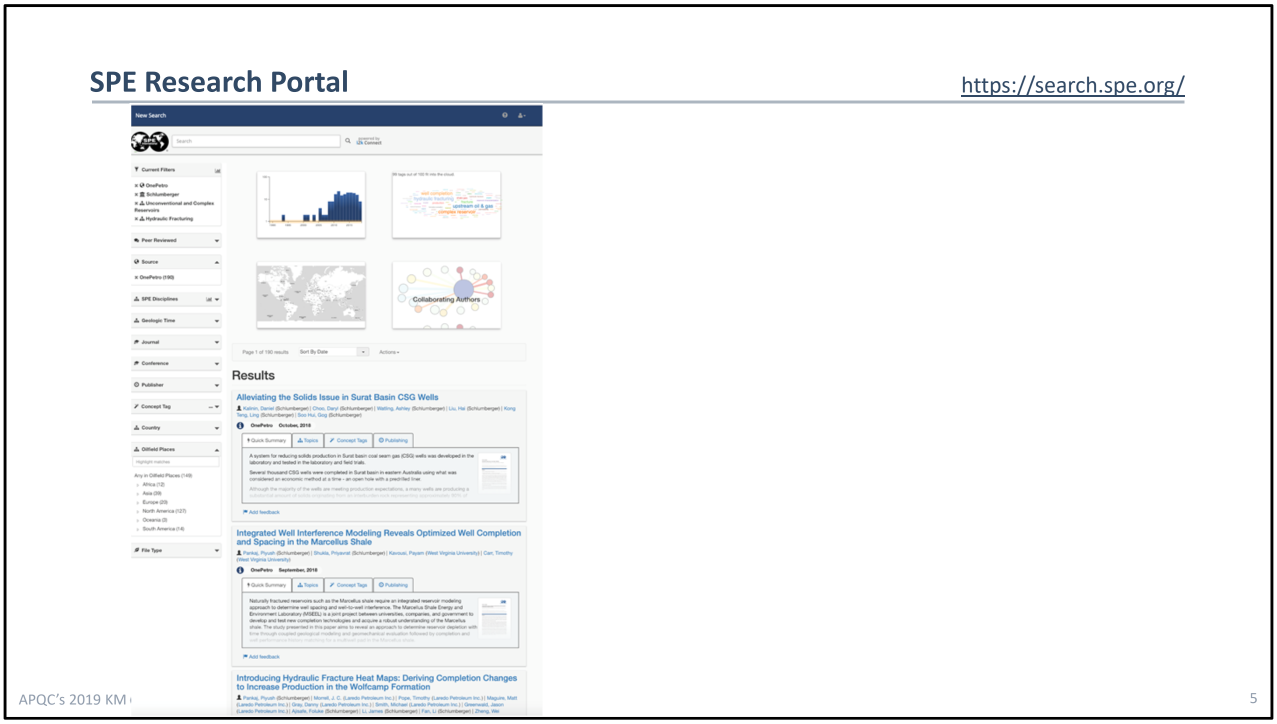
- Over 200,000 distinct items on SPE's web sites
- Content doubling every 10-11 years, since the 1950's
- Many new channels

- Finding information and subject matter experts is hard
- Sheer amount of content obscures industry trends



- Solution: Research Portal to improve findability and consistency
- Partner with i2k Connect – deep experience in both AI and E&P

From a Knowledge Management point of view, the Research Portal is intended achieve two goals: connecting people with the information and the subject matter experts needed to solve problems, and improving cross-industry knowledge flow.



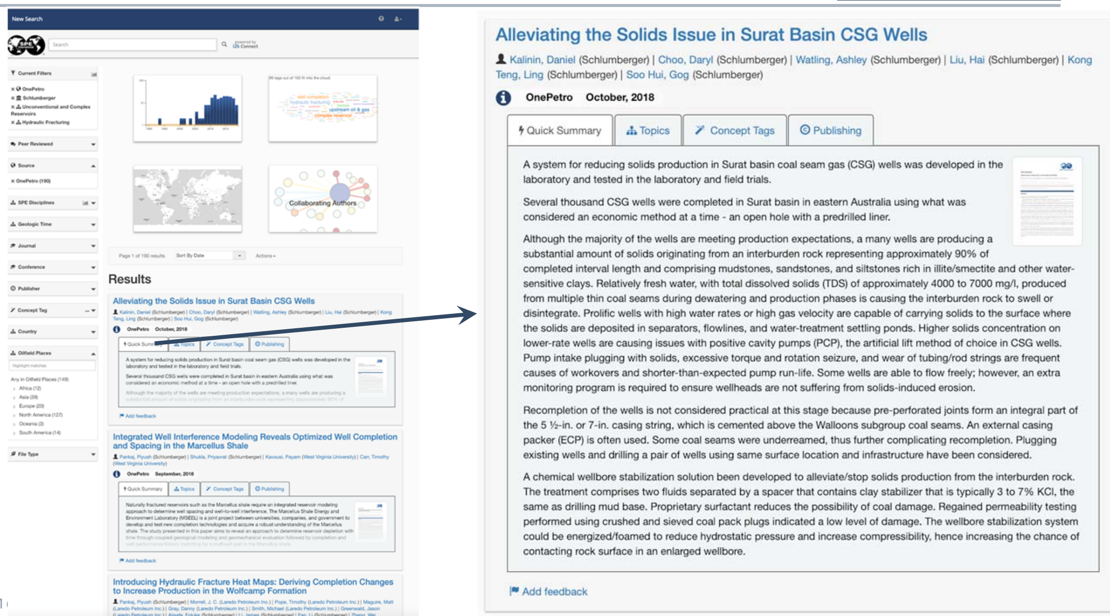
SPE Research Portal: supported by artificial intelligence (AI). Integrates SME knowledge with AI natural language processing and machine learning. Automatically enriches documents by classifying them into relevant taxonomies, geotagging oil fields, and extracting key concepts, authors, and institutions. These enrichments enable SPE members to zero in on the relevant information from all SPE channels and to graphically analyze timeframes, geography, related concepts and cross industry collaboration (using social network analysis).

https://search.spe.org/i2kweb/SPE/search?filters=taxnodes:SPE%20Disciplines%7CReservoir%20Description%20and%20Dynamics%7CUnconventional%20and%20Complex%20Reservoirs%40%40taxnodes:SPE%20Disciplines%7CWell%20Completion%7CHydraulic%20Fracturing%40%40store:OnePetro%40%40institutions:Schlumberger

Let's start with the information displayed by the Portal for a single item.

SPE Research Portal

https://search.spe.org/



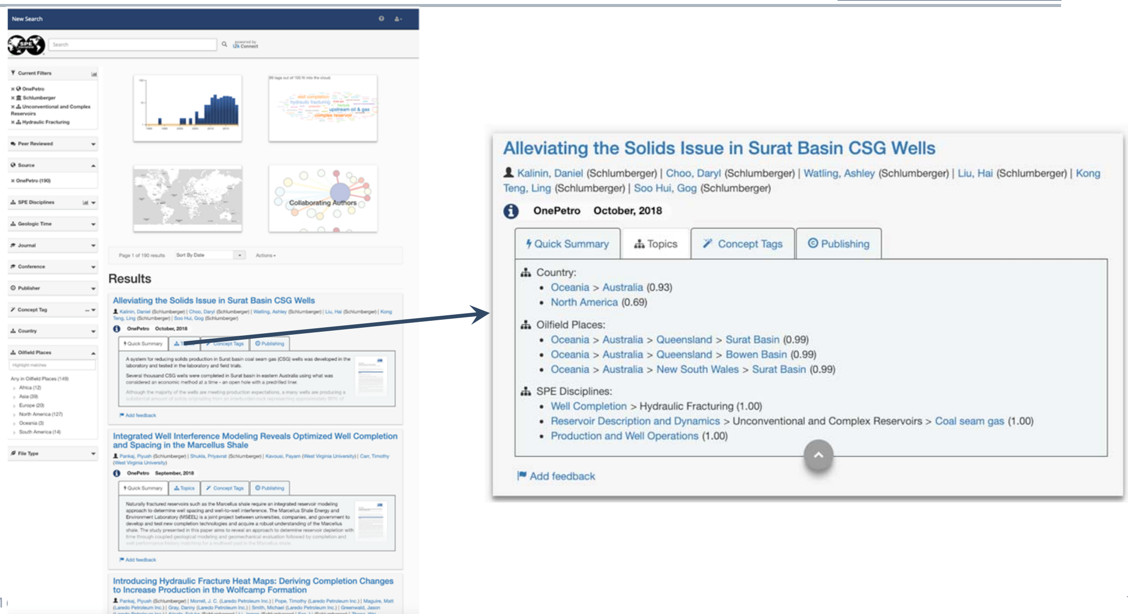
APQC's 2019 KM 6

https://search.spe.org/i2kweb/SPE/search?filters=taxnodes:SPE%20Disciplines%7CReservoir%20Description%20and%20Dynamics%7CUnconventional%20and%20Complex%20Reservoirs%40%40taxnodes:SPE%20Disciplines%7CWell%20Completion%7CHydraulic%20Fracturing%40%40store:OnePetro%40%40institutions:Schlumberger

Quick Summary / Abstract / Snippet

SPE Research Portal

https://search.spe.org/



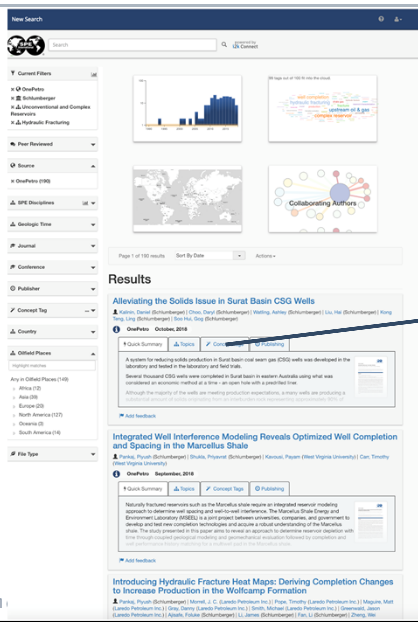
APQC's 2019 KM 7

https://search.spe.org/i2kweb/SPE/search?filters=taxnodes:SPE%20Disciplines%7CReservoir%20Description%20and%20Dynamics%7CUnconventional%20and%20Complex%20Reservoirs%40%40taxnodes:SPE%20Disciplines%7CWell%20Completion%7CHydraulic%20Fracturing%40%40store:OnePetro%40%40institutions:Schlumberger

Classifications – from three of the four taxonomies used in the SPE Research Portal. Out of the box, the i2k AI Platform classifies into 17 taxonomies.

SPE Research Portal

https://search.spe.org/

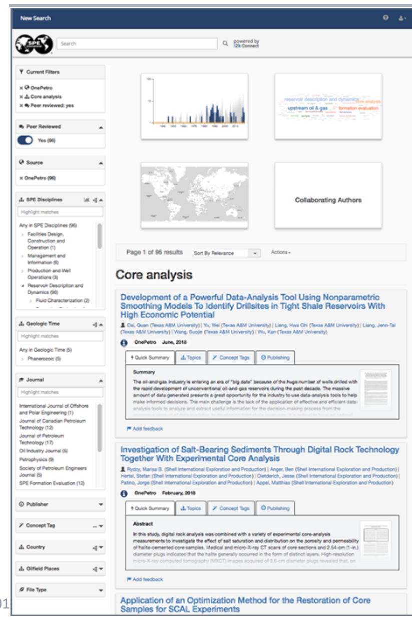


https://search.spe.org/

https://search.spe.org/i2kweb/SPE/search?filters=taxnodes:SPE%20Disciplines%7CReservoir%20Description%20and%20Dynamics%7CUnconventional%20and%20Complex%20Reservoirs%40%40taxnodes:SPE%20Disciplines%7CWell%20Completion%7CHydraulic%20Fracturing%40%40store:OnePetro%40%40institutions:Schlumberger

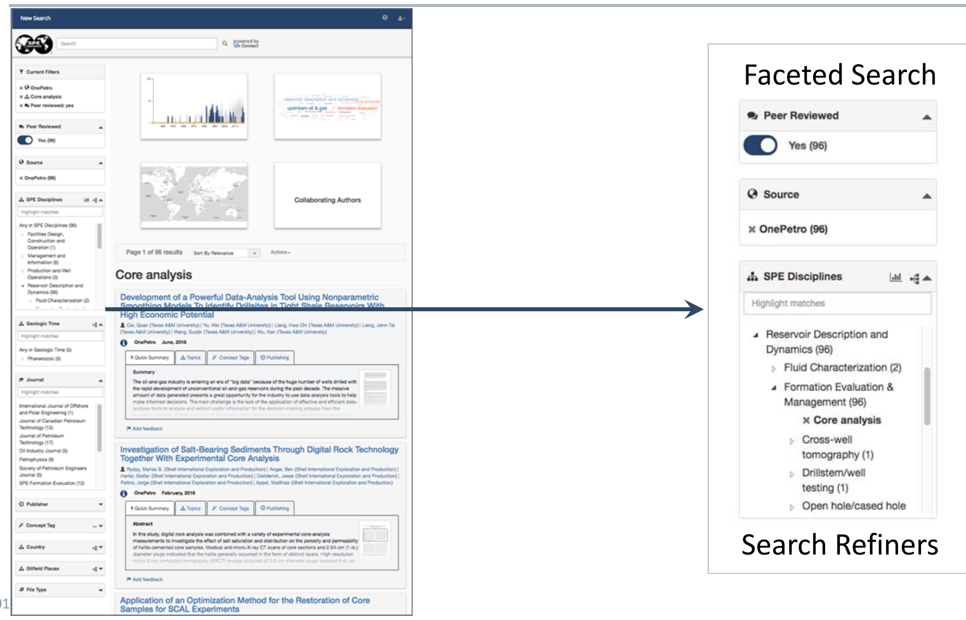
Concept Tags – key words and phrases

SPE Research Portal



Let's now discuss how to find information and subject matter experts, plus how to analyze information.

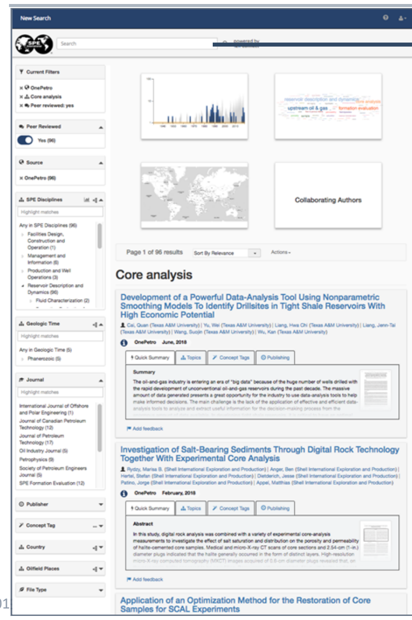
SPE Research Portal



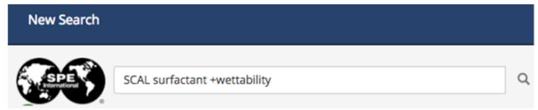
The search refiners for faceted search appear on the left.

https://search.spe.org/i2kweb/SPE/search?filters=peer-reviewed%3Atrue%40%40store%3AOnePetro%40%40taxnodes%3ASPE+Disciplines%7CReservoir+Description+and+Dynamics%7CFormation+Evaluation+%26+Management%7CCore+analysis

SPE Research Portal

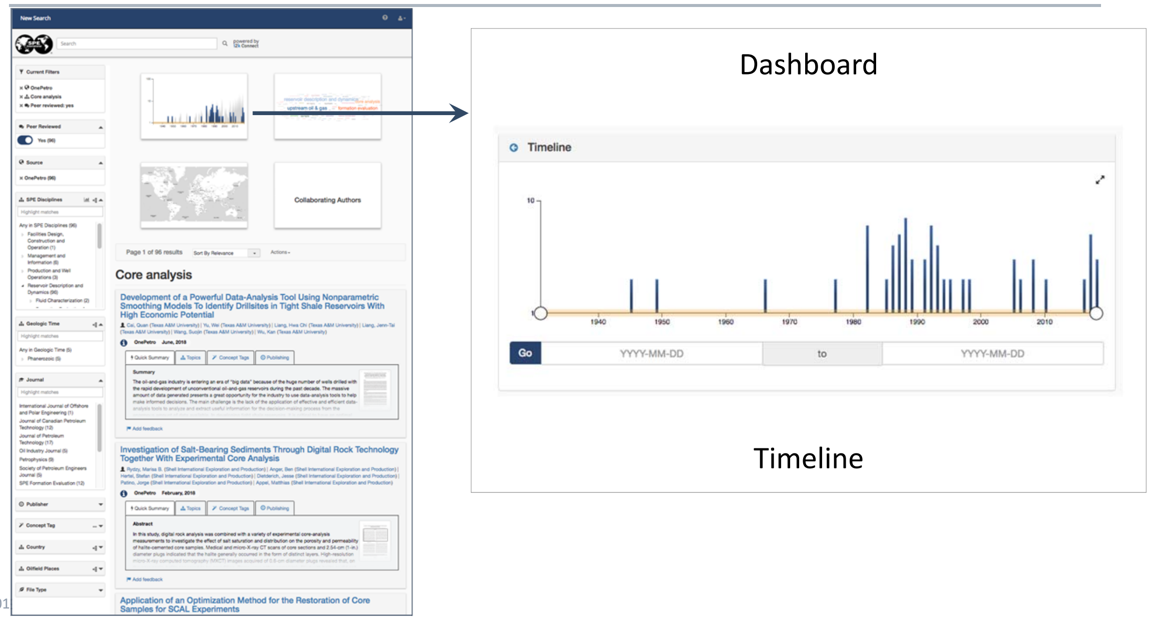


Text Search



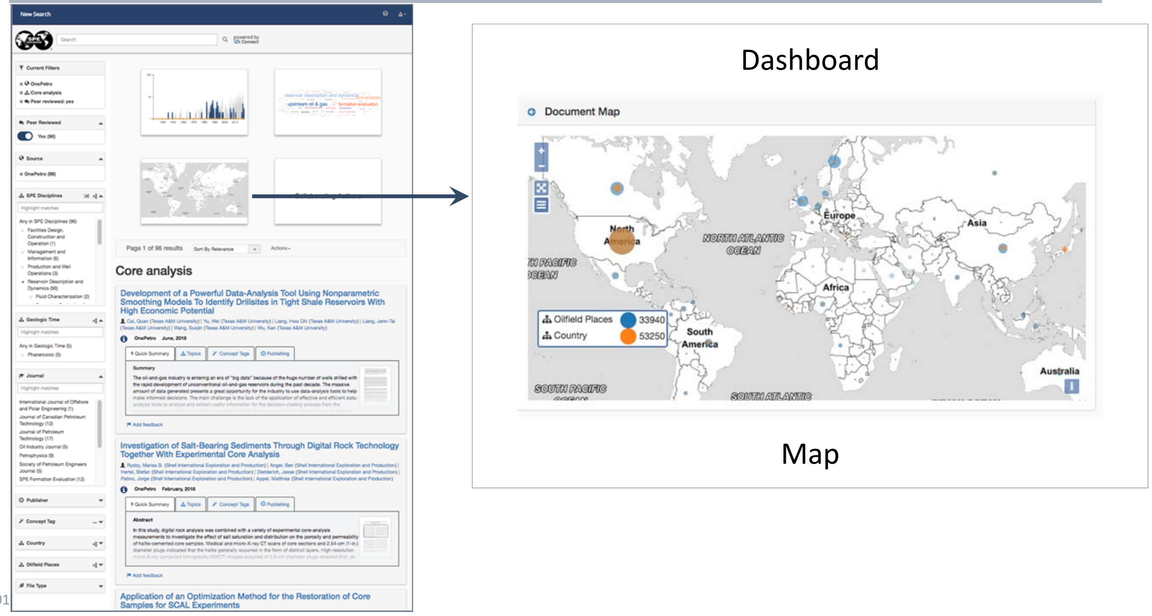
https://search.spe.org/i2kweb/SPE/search?filters=peer-reviewed%3Atrue%40%40store%3AOnePetro%40%40taxnodes%3ASPE+Disciplines%7CReservoir+Description+and+Dynamics%7CFormation+Evaluation+%26+Management%7CCore+analysis

SPE Research Portal

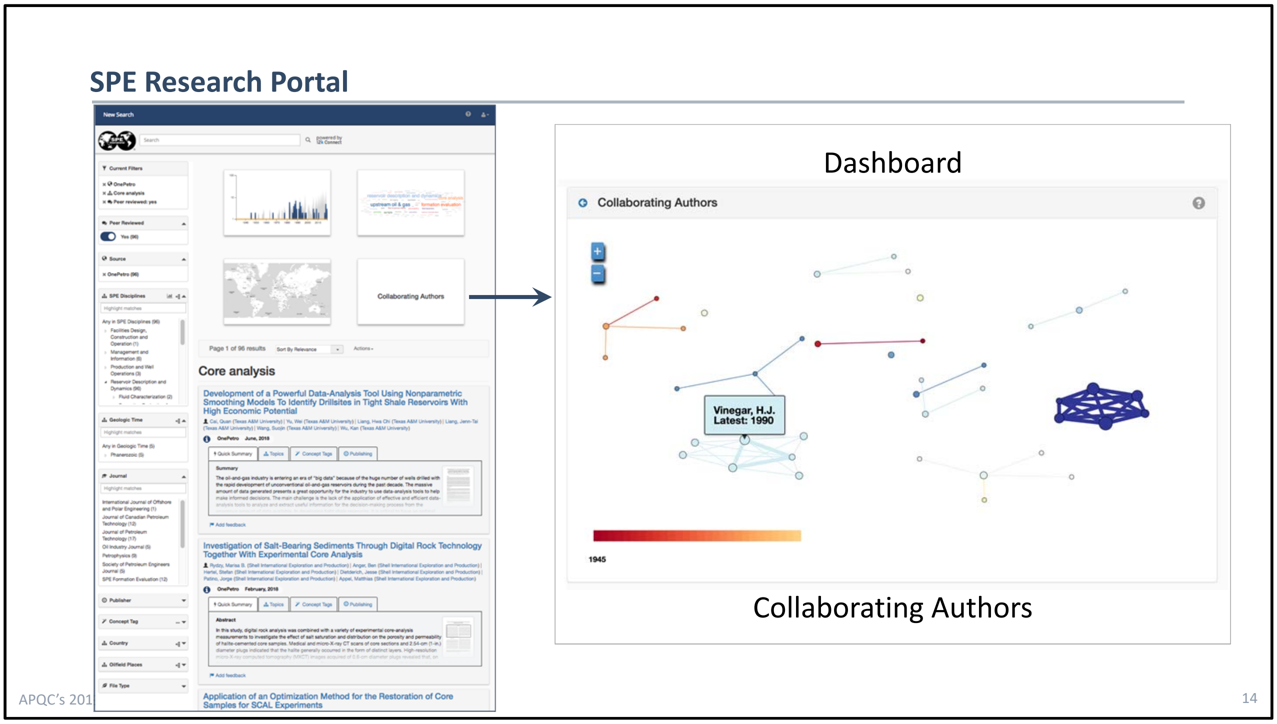


https://search.spe.org/i2kweb/SPE/search?filters=peer-reviewed%3Atrue%40%40store%3AOnePetro%40%40taxnodes%3ASPE+Disciplines%7CReservoir+Description+and+Dynamics%7CFormation+Evaluation+%26+Management%7CCore+analysis

SPE Research Portal



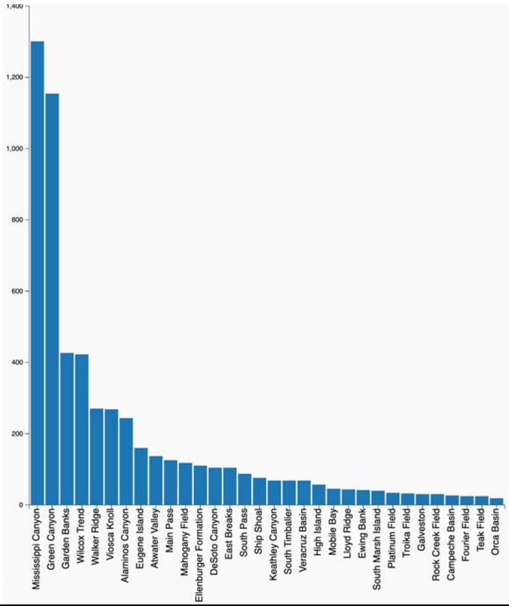
https://search.spe.org/i2kweb/SPE/search?filters=peer-reviewed%3Atrue%40%40store%3AOnePetro%40%40taxnodes%3ASPE+Disciplines%7CReservoir+Description+and+Dynamics%7CFormation+Evaluation+%26+Management%7CCore+analysis



This panel shows collaborations among authors of papers. A simple “Social Network Analysis” grapy. The size of a bubble is an indicator of how many papers the person has authored. The connections show authors who have collaborated. The thickness is an indicator of how many papers they have co-authored. You can click on a bubble to reveal the author and go to his/her papers. You can see a few individual authors, plus clusters of authors. The very tight cluster in dark blue is a group at Texas A&M.
This panel is particularly useful for identifying subject matter experts.

https://search.spe.org/i2kweb/SPE/search?filters=peer-reviewed%3Atrue%40%40store%3AOnePetro%40%40taxnodes%3ASPE+Disciplines%7CReservoir+Description+and+Dynamics%7CFormation+Evaluation+%26+Management%7CCore+analysis

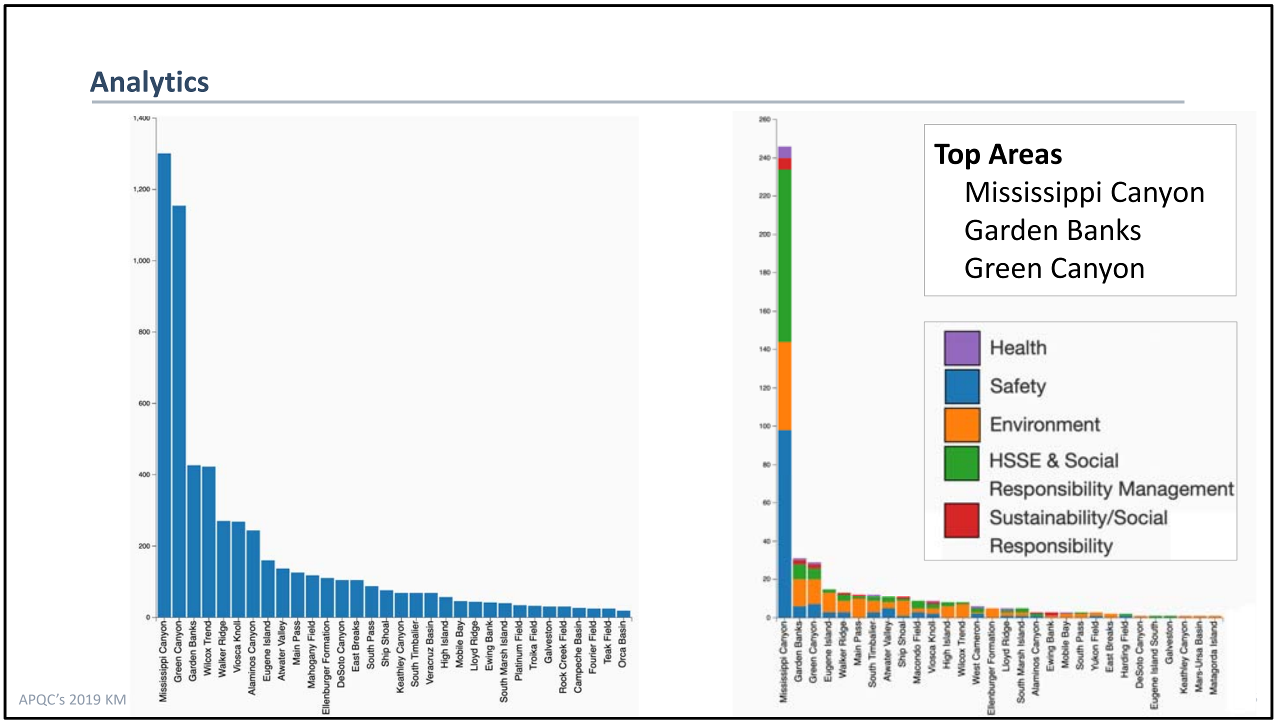
Analytics



Top Areas
- Mississippi Canyon
- Green Canyon
- Garden Banks

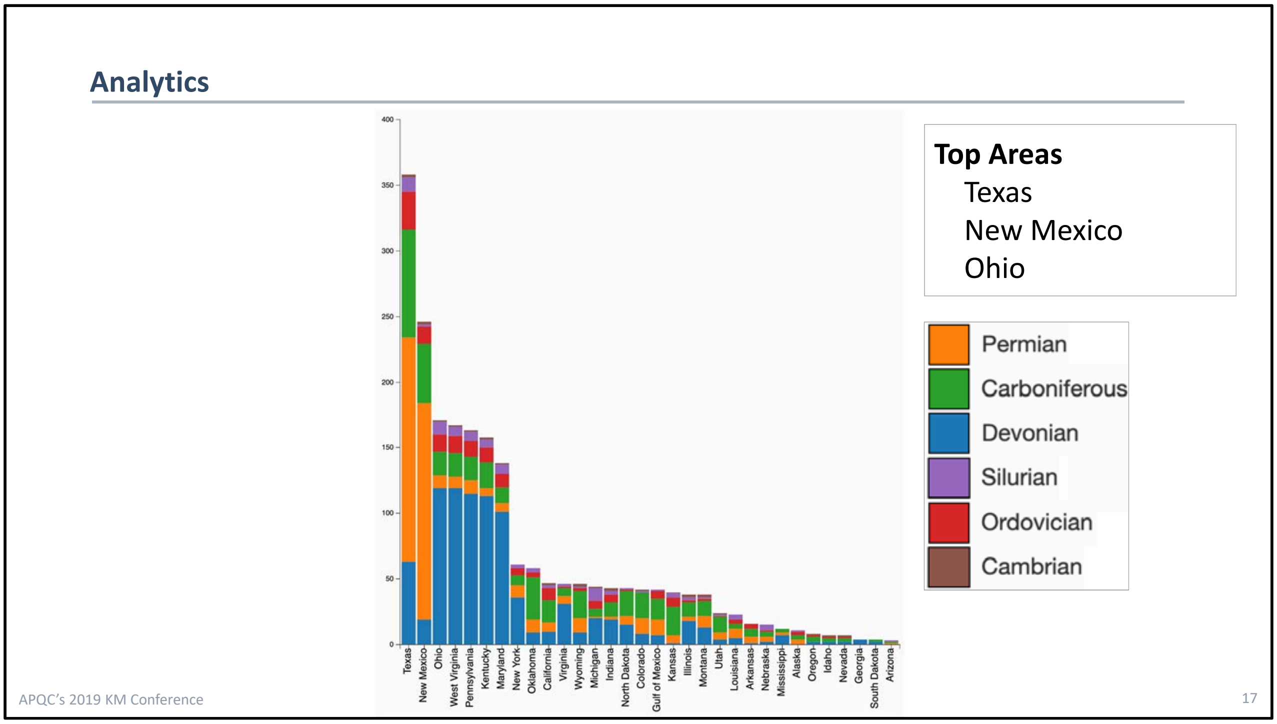
APQC's 2019 KM 15

All OnePetro content related to the Gulf of Mexico.
https://search.spe.org/i2kweb/SPE/search?filters=store:OnePetro%40%40taxnodes:Oilfield%20Places%7CNorth%20America%7CUnited%20States%7CGulf%20of%20Mexico



First graph: All OnePetro content related to the Gulf of Mexico.
https://search.spe.org/i2kweb/SPE/search?filters=store:OnePetro%40%40taxnodes:Oilfield%20Places%7CNorth%20America%7CUnited%20States%7CGulf%20of%20Mexico

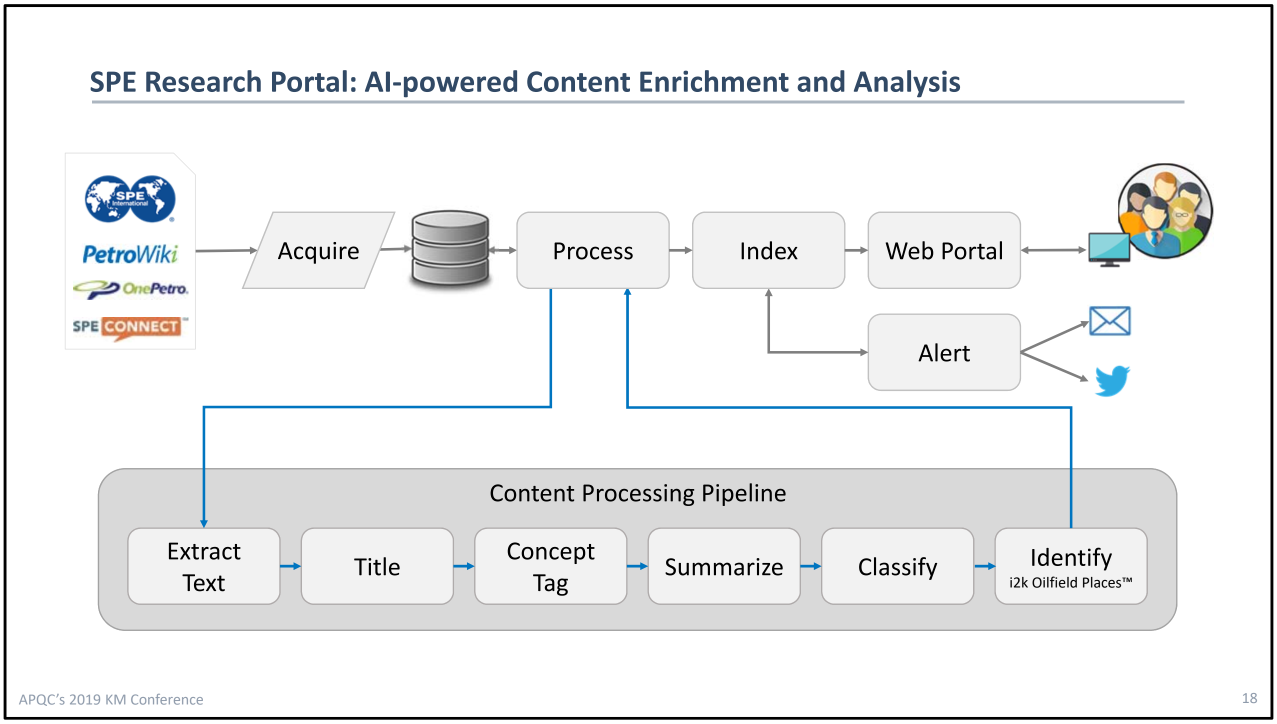
Second graph. Filter by Health, Safety, Security, Environment, and Social Responsibility. Note Mississippi Canyon dominates, due to Macondo.



US States by Paleozoic era.

https://search.spe.org/i2kweb/SPE/search?filters=store:OnePetro%40%40taxnodes:Oilfield%20Places%7CNorth%20America%7CUnited%20States

Permian (orange) in Texas/New Mexico. Devonian (blue) – Marcellus Shale.

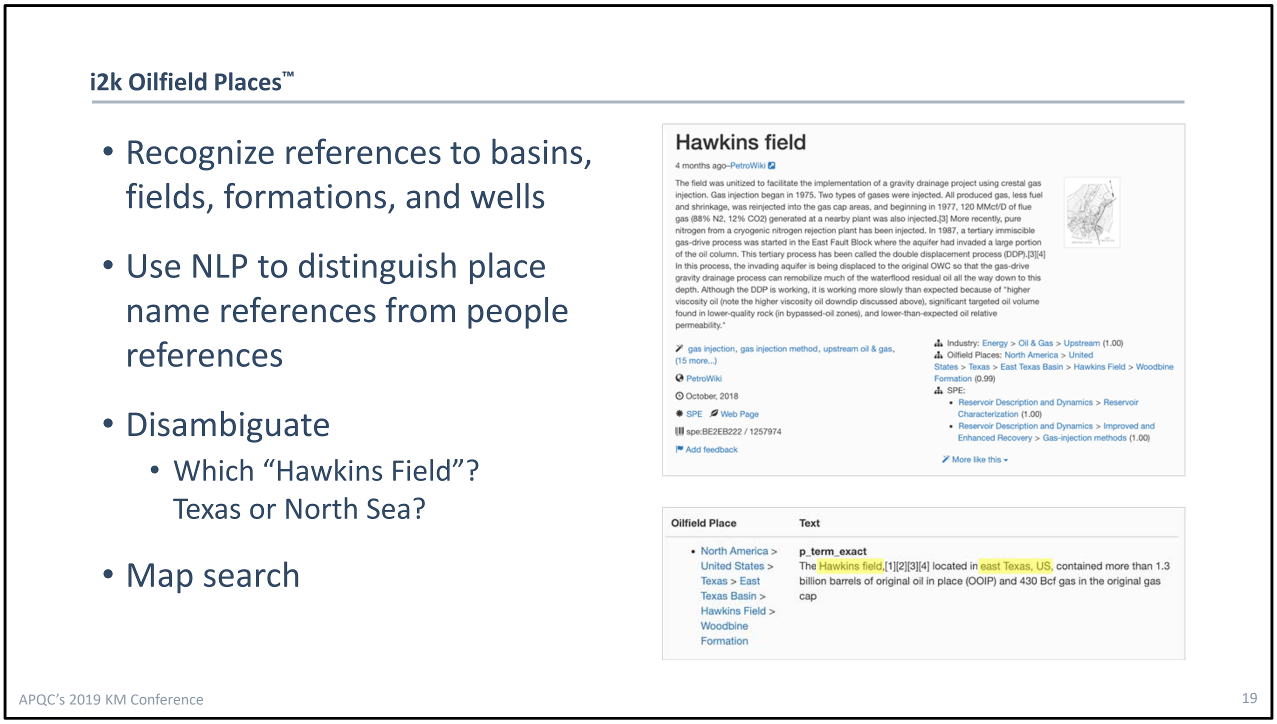


Collects content from OnePetro, PetroWiki, SPE.org, and in the future, SPEConnect.
Enriches content by adding metadata:

- Titles: as opposed to filenames
- Concept tags: key words and phrases
- Summaries, authors, organizations, dates
- Classifications (SPE Disciplines, Geologic Time, ...)
- Identified entities: i2k Oilfield Places™

Indexes content to allow search by text, classification, location, author, …

Supports simple analytics to expose trends, distributions and correlations.



Disambiguate names

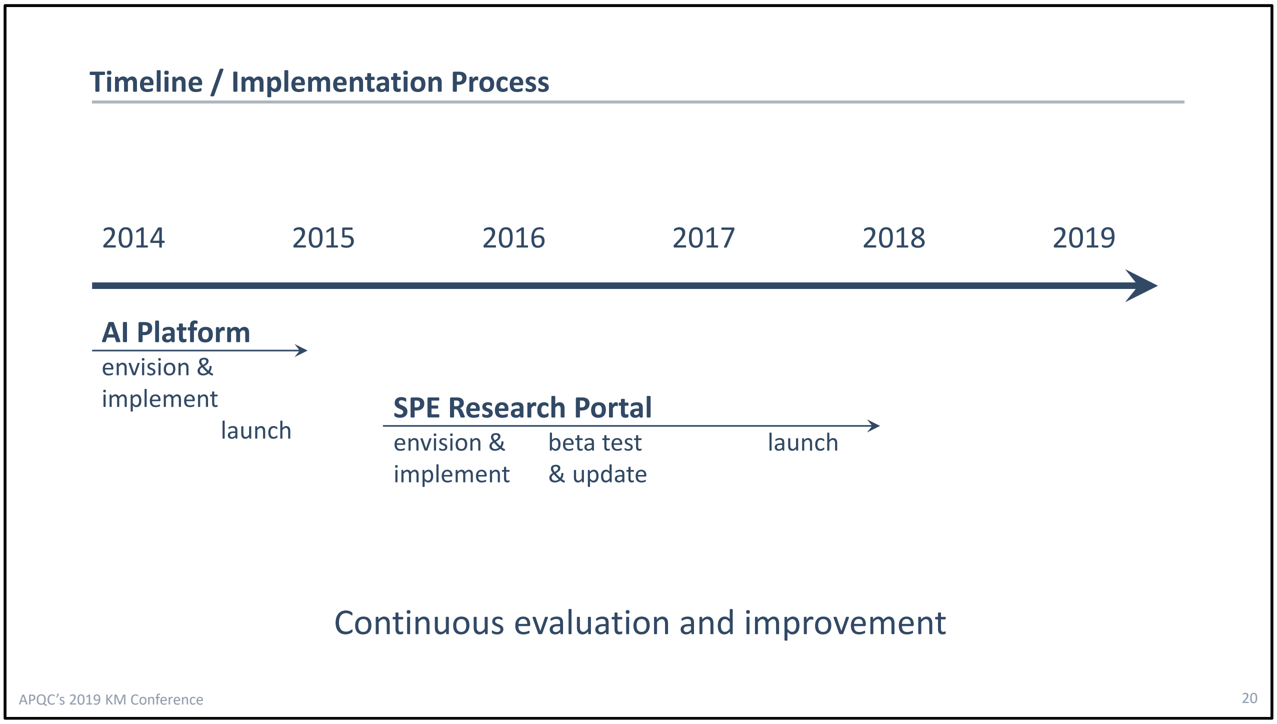
Use geopolitical and other "place" clues in the original content

Weigh evidence for best match

So… if we see "Good Omen" and Hawkins in proximity, we'll conclude Texas, because both are in the East Texas Basin, and "Good Omen" is unambiguous.

Map interface: Correlate with labelled spatial GIS features on maps

https://search.spe.org/i2kweb/SPE/doc/petrowiki:7D41618D/



It took approximately two weeks for the first processing of SPE content in 2015.
We had a working Research Portal from that point on.
Since then we have made a number of design improvements based on tester and end-user feedback (e.g., adding the dashboard).

Knowledge base improvements are made on a continuous basis.

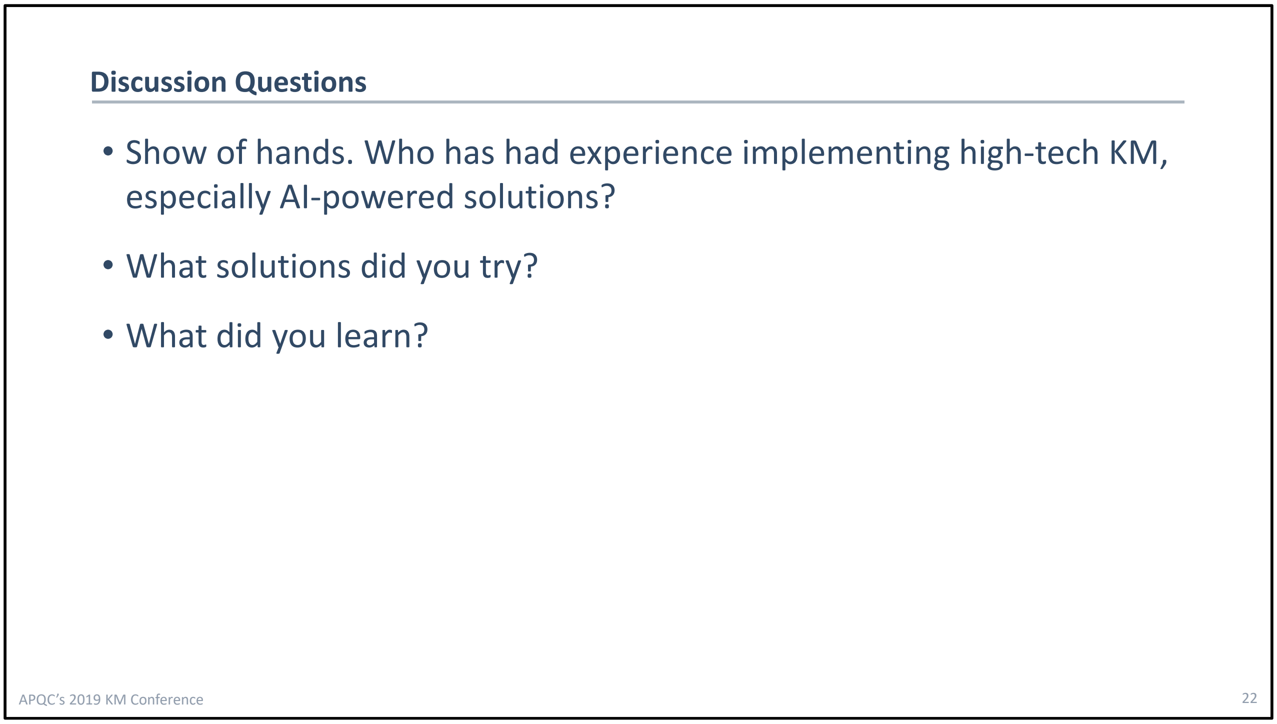
We have also used the same platform for other knowledge bases and industries. This is a benefit of a platform approach as opposed to a one-off solution. We also benefit from synergy – what is learned in one implementation carries over to the others. There is also synergy within a single industry. There is considerable overlap between the knowledge bases for one company in an industry and the knowledge bases for a different company in the same industry.

More recently, we have added capabilities for table identification and extraction of structured data from documents.

Results – Measures of Success

- Google Analytics
  - Average number of non bot hits to new Research Portal is 26,055 per month.
  - Average number of regular SPE search hits is 3,797 per month.
  - Average hold time on new Research Portal after a search and number of pages read is double regular SPE search average.
- Anecdotes

Average hold time on the new research portal after a search and the number of pages read is double the normal average.
That could imply that people are finding what they want to read though we have no proof of that.



This slide is for the group discussion part of the presentation.

Our responses to survey questions posed by APQC.

Track: High-Tech Solutions to Knowledge Problems
Why do so many new technologies end up disappointing the KM programs that implement them?

- They don't provide enough power/functionality that meets the needs of users.
- They demand too large a change to users' existing operating procedures / ways of working.
- They do not incorporate sufficient subject matter expertise / knowledge of users' industries/companies/jobs to be helpful.
- They cannot explain the help they provide / the decisions they take.

How can great KM programs ensure their high-tech solutions are relevant to users' needs?

- Users are directly involved in establishing the requirements, success criteria, project plans, communications plans, and system evaluation/improvement.
- Benchmark to understand what has been successful in similar companies - critical success factors, best practices and lessons learned.
- A realistic budget is agreed in advance.
- Governance of is key and technology adds another layer of complexity to the governance of KM projects. In addition to the traditional business/operational partners, the IT department is an essential partner.

What is the most surprising benefit or capability of new KM technologies that people may not be aware of?

- AI technology can read and analyze documents that go well beyond keyword analysis to assist in connecting people to the information they need. This includes identifying data and answers to questions.
- AI technology can automate the social network analysis that is key to connecting people to the people who have the experience and knowledge to answer questions.

How can KM teams separate hype from reality when evaluating new tools and applications?

- Benchmark to understand what has been successful in similar companies - critical success factors, best practices and lessons learned.
- Partner with technology providers to understand how to get the most out of existing products and adapt existing products to meet your specific needs. This is especially true for AI products since at this immature stage of the market, most products require significant customer input to produce high-quality results.

Our Lessons Learned from Implementing AI-powered Solutions

- AI systems will enable knowledge workers to focus on high-value activities by automating the repetitive, manual tasks of finding, extracting, and normalizing data in documents
- More likely automated intelligent assistants than “set and forget”
- AI is not a DIY activity – budget for it – partner with or hire AI experts
- Domain Knowledge is critical
- Machine Learning can accelerate knowledge acquisition but is not magic
- Continuous evaluation and improvement after deployment are essential
- There is no substitute for people and processes

Consistency and repeatability are important for findability – and the current manual approach is neither consistent nor repeatable.
AI systems will free up ... This is another way of saying that manual information governance is not sustainable. Companies have been depending on individual workers to manually tag documents so that they will be findable. And they don’t do it. As a result, as one Operator VP told us: “Our information is either lost in the cloud today, or on its way to being lost.” Employees resist being burdened with metadata entry, and they don’t do it very well.
So what have we learned about implementing AI-powered solutions?
Intelligent Assistants are more likely than completely automated solutions. So ... As with any automation, minimize extraneous changes to existing human-powered workflows. Consider explainability, handoffs, etc.
AI is not a DIY activity ... Suggestions: Work with suppliers that know your industry, as well as AI. Being adept at AI is not enough. Buy a subscription to a service that improves over time by leveraging the learnings across the industry.
Domain Knowledge is critical – one of the earliest learnings in applied AI. To think about what knowledge to incorporate into AI systems, consider how you would train a team of colleagues to do the work.
Knowledge Acquisition Bottleneck: Manual knowledge acquisition is hard. Machine Learning is a way around the bottleneck ... but there is no magic. We discussed this earlier, but the point bears repeating. Machine Learning is widely applied to Big Data, but behind the scenes there is a wealth of human prep work. See: Reid G. Smith and Joshua Eckroth. Building AI Applications: Yesterday, Today, and Tomorrow. AI Magazine 38(1): 6-22, Spring 2017. From MIT Professor Howie Shrobe: “... when you look closer at successful statistical approaches, a lot of the success is in the choice of features to attend to or other similar ways of conveying human insight to the technique ...” In addition, you won’t be able to explain results, you will need many training examples … and who does the work to create the examples or can it be automated?
Continuous evaluation and improvement after deployment. Things won’t be perfect out-of-the-box.
Finally, there is no magic in AI, including machine learning. Humans guidance is essential. And as in all aspects of the Oil & Gas industry, robust processes are required to guide the humans. Two examples: 1) You will be implementing or introducing new software. History has shown how tricky that can be without processes for requirements gathering, change management, testing. 2) Taxonomy design must be supported by a process that ensures the taxonomies will be useful to the users. Otherwise, they may do more harm than good.

Acknowledgments / Thank You / Questions

- Thanks to the SPE staff and directors who have provided support, feedback, and encouragement.
- This work has been supported in part by the National Science Foundation Small Business Innovation Research Program, Award #1415757 and #1534798.