

The Comparison of the Petroleum Context for the Term "Digital transformation" from OnePetro with General Technician Context from Scopus

Objective: to obtain key phrases from OnePetro on "Digital Transformation" topic (petroleum context) and compare them with key phrases from Scopus (more general context). In Scopus, we try to limit context by technical topics, see Scopus query in the text below. The purpose of such comparison is to find nonordinary items in general context which could be useful for petroleum context.

Emotional remark: I was deeply impressed when text mining of titles and abstracts from Scopus prompted the phrase "**cyber-physical systems data analytics**" (Table 2) and then in dimensions.ai the query "'cyber-physical systems" AND analytic' suggested **31** publications in 2010 and **2030** in 2018!

Common remark for tables:

OnePetro query: search for "Digital Transformation" AND conference papers AND since 2015; 117 results

OnePetro: text based on 117 titles and abstracts

Scopus query: (TITLE-ABS-KEY ("digital transformation" OR "digital oilfield" OR "digital technolog*" OR "digital twin" OR "digital solution*" OR "digital system*" OR "digital environment*" OR "digital data technolog*" OR "digital business" OR "digital compan* data") OR TITLE-ABS-KEY ("digital strategy" OR "digital information" OR "digital design" OR "digital monitoring system" OR "drilling process digitalization")) AND PUBYEAR > 2014 AND (LIMIT-TO (SRCTYPE , "j")) AND (EXCLUDE (SUBJAREA , "SOCJ") OR EXCLUDE (SUBJAREA , "ARTS") OR EXCLUDE (SUBJAREA , "PSYC")) AND (LIMIT-TO (LANGUAGE , "English")) ; 4,128 results

Scopus: text based on 2000 most cited titles and abstracts

Score: as calculated by "textextract" from KH Coder 3a16f;

1299 stop words were removed

$A \cap B$: column A and B intersection

Table 1. Comparison of top 40 Key Phrases from files: OnePetro-Words clusters by textextract.tsv and KeyPhrases_scopus-titles-abstracts.tsv

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score	$A \cap B$
gas industry	1417.818	digital technologies	343870,8974	information technology
digital transformation	1391.907	digital technology	338816,8225	digital technologies
big data	943.046	digital information	232944,8815	digital technology
real-time data	800.874	digital systems	120710,1472	digital transformation
digital oilfield	539.413	digital design	95945,00736	digital twin
digital technologies	523.325	digital transformation	77639,2358	big data
data analytics	518.631	digital system	75524,87342	
data management	452.683	digital health	38912,09524	
reservoir management	410.463	digital twin	32141,67109	
digital twin	381.208	big data	30190,20027	
oil production	359.831	social media	26103,28678	
real time data	344.820	digital data	24023,69785	
digital technology	337.430	digital environment	18007,624	
big data analytics	335.183	digital media	13752,42185	
oil gas industry	301.470	health information	11803,00606	
production operations	287.011	case study	11392,22671	
real time	256.072	new digital technologies	10706,9794	
gas production	244.927	new technologies	10607,15166	
new data	240.758	digital economy	10387,71659	
oil gas	232.754	future research	9940,715337	
offshore oil	226.478	health care	9890,2133	
drilling operations	220.204	digital business models	9708,500056	
gas operators	196.916	information technology	9399,838457	
data management analytics	195.448	digital health technologies	8555,347143	
gas companies	193.184	digital innovation	8368,096342	
historical data	188.037	digital age	8243,554319	

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score	A ∩ B
data integration	175.765	information systems	7790,295565	
energy industry	170.009	digital twins	7365,448856	
process safety	164.401	digital designs	6842,334231	
production data	157.634	digital pathology	6357,321625	
advanced analytics	155.999	new digital technology	6276,922211	
information technology	155.626	data analysis	6105,345135	
data security	149.077	mental health	5852,316933	
data platform	145.810	digital solutions	5513,071089	
data science	138.732	digital platform	5503,810715	
quality of data	138.732	simulation results	5490,058167	
oil field	135.111	digital platforms	5481,494002	
industrial data platform	134.697	digital system design	5383,959729	
production optimization	133.834	digital disruption	5249,458731	
monitoring process	129.460	digital devices	4974,868297	

If we take all 500 key phrases from OnePetro and comprise them with 48667 key phrases from Scopus, we get the following intersection (119 terms) which are enough for domain description:

(file: A and B for all terms.txt)

real time data; technical challenges; software platform; data transmission; monitoring process; time consuming; historical data; data analytics; case studies; real-time information; digital technology; digital design; supply chain; energy consumption; huge amount of data; available data; production data; high value; digital information; energy industry; digital transformation process; predictive analytics; data availability; artificial intelligence; new system; cyber security; asset management; industrial companies; full potential; digital strategy; digital network; control system; different data; predictive maintenance; value chain; digital technologies; unmanned vehicles; multiple data sources; amount of data; digitalization process; reference data; digital systems; data acquisition; new technology; new data; different systems; automation system; life extension; new approach; big data; management systems; new business models; data sets; new opportunities; business models; low cost; detailed information; operational excellence; case study; advanced analytics; digital solutions; same time; digital technology adoption; cost savings; data repositories; new technologies; information technology; digital capabilities; project management; data science; real-time communication; production operations; high efficiency; quality of data; real time; maritime industry; data access; information security; artificial intelligence ai; such technology; digital business; value of data; oil gas; digital service; operational issues; daily operations; use of information technology; industrial internet of things; digital twin; oil companies; visual data; data governance; data processing; digital transformation strategy; digital environment; data security; digital transformation; new ways; design process; oil price; data set; business decisions; blockchain technology; real-time data; smart operations; decision process; decision making; relevant data; lack of data; big data analytics; data management; business processes; digital monitoring system; first time; set of data; new information; measured data; sensor data; production system

Table 2. Comparison top 30 Key Phrases of "'All Key Phrases AND 'analytic' filter" from files: 'OnePetro-Words clusters by textextract analytics.csv' and 'KeyPhrases_scopus-titles-abstracts-analytic.tsv'

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score	A ∩ B
data analytics	518.631	big data analytics	2214.291	data analytics
big data analytics	335.183	data analytics	1943.972	big data analytics
data management analytics	195.448	business impacts of data analytics	252.070	
advanced analytics	155.999	use of analytics	239.603	
advanced data analytics	112.387	use of data analytics	235.031	
big data analytics techniques	73.333	big data analytics capabilities	202.156	
real-time drilling downhole tools data analytics	71.904	use of cloud systems data analytics	187.533	
predictive analytics	69.757	use of big data analytics	182.159	
diagnostic analytics	51.576	applications of data analytics	174.555	
predictive analysis	47.916	business analytics ecosystems	159.971	
advanced big data analytics	47.212	cyber-physical systems data analytics	148.705	
integration of big data analytics	44.551	big data analytics industry	142.637	
analytics system	41.367	data analytics adoption	137.055	
artificial intelligence data analytics	38.907	business analytics	131.025	
analysis process	36.373	adoption of big data analytics	121.557	
big data storage analytics	35.840	potential use of analytics	117.617	

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score	A ∩ B
real-time analytics	35.163	future of process analytics	114.952	
several big data analytics	34.278	intersection of big data analytics	111.591	
big data analytics framework	33.623	analytical design control methodology	109.532	
data storage data analytics reliability improvement	33.517	supply chain control analytics	109.426	
petrophysical big data analysis	31.161	industrial internet of things cloud computing big data analytics	106.825	
analytics models	28.098	analytical methods	102.576	
		analytical technology	101.397	
		accounting big data analytics business analytics	97.753	
		quality analytics	94.902	
		adaptation of data analytics	94.102	
		potential impacts of big data analytics	93.869	
		analytical applications	93.016	
		analytical techniques of data	92.012	
		visual analytics	90.050	

Remark: Common topic: "Digital Transformation AND Analytic" (Scopus) is more **business** and **cloud** oriented compare to petroleum domain. cyber-physical systems data analytics: in OnePetro search for "cyber-physical systems data analytics" has returned 0 results; in OnePetro search for "cyber-physical systems" has returned 22 results; search for "cyber-physical" has returned 35 results.

Hypothesis: "cyber-physical systems data analytics" could be an attractive emergence topic for petroleum domain based on experience in other industries. **Emphasis on analytics** for cyber-physical systems

A **cyber-physical** (also styled **cyberphysical**) **system (CPS)** is a **mechanism** that is controlled or monitored by computer-based algorithms, tightly integrated with the Internet and its users. In cyber-physical systems, physical and software components are deeply intertwined, each operating on different **spatial and temporal scales**, exhibiting multiple and distinct behavioral modalities, and interacting with each other in a lot of ways that change with context.

https://en.wikipedia.org/wiki/Cyber-physical_system

Query "'cyber-physical systems' analytics' in **dimensions.ai** return:

PUBLICATIONS YEAR: 2019 2,162; 2018 3,020; 2017 1,885; 2016 1,103; 2015 622; 2014 337; 2013 193; 2012 121; 2011 77; 2010 31

The nine years grow (2010 - 2018) is: 3020/31=**97.4** , compare **2** times for common topics

Table 3. Comparison top 40 Key Phrases of "All Key Phrases AND "data" filter" from files: 'OnePetro-Words clusters by textextract data.csv' and 'KeyPhrases_scopus-titles-abstracts-data.tsv'

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score	A ∩ B
big data	943.046	big data	30190.200	new data
real-time data	800.874	digital data	24023.698	data analytics
data analytics	518.631	data analysis	6105.345	real-time data
data management	452.683	data collection	3780.141	data management
real time data	344.820	data management	3122.727	relevant data
big data analytics	335.183	health data	2472.536	big data analytics
new data	240.758	big data analytics	2214.291	data science
data management analytics	195.448	new data	2145.124	big data
historical data	188.037	patient data	2126.394	
data integration	175.765	data analytics	1943.972	
production data	157.634	big data analysis	1858.165	
data security	149.077	data quality	1589.793	
data platform	145.810	such data	1589.630	
data science	138.732	real-time data	1485.281	
quality of data	138.732	research data	1341.059	
industrial data platform	134.697	empirical data	1286.674	
downhole data	121.551	data processing	1167.347	

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score	A ∩ B
oil gas data	116.473	qualitative data	1144.165	
advanced data analytics	112.387	digital design data	1082.704	
big data service	100.611	big data applications	1030.144	
real-time data environment	97.570	data acquisition	981.197	
big data environment	95.177	amount of data	888.947	
value of data	89.035	experimental data	838.790	
big data management	86.067	literature data	815.234	
data access	79.372	data information	809.665	
offshore data	79.372	environmental data	802.515	
integrity data	77.368	data sources	775.663	
effective data	73.864	clinical data	774.172	
big data analytics techniques	73.333	sensor data	762.203	
relevant data	73.170	data storage	736.038	
real-time drilling downhole tools data analytics	71.904	data science	721.659	
subsea data	71.391	limited data	700.205	
transient data	68.525	electronic data	691.517	
digital data technologies	68.364	data models	688.773	
key data	67.428	research design data	684.243	
data management results	65.620	data set	680.903	
digital companies data	63.112	digital transformation of data	662.306	
surveillance data	62.957	data mining	628.225	
technology data standards	61.869	sharing of data	623.622	
different data sources	60.936	relevant data	612.533	

Definition: Qualitative data consist of words, observations, pictures, and symbols. Qualitative Data Analysis (QDA) refers to the processes and procedures that are used to analyze the data and provide some level or explanation, understanding, or interpretation. Qualitative data analysis typically occurs simultaneously with the data collection. Therefore, meaning and understanding often develop slowly over time in a non-linear fashion as the project progresses.

https://cirt.gcu.edu/research/developmentresources/research_ready/qualitative/analyzing_data

Remark: The following topics could have additional interest for petroleum domain: sensor data; digital transformation of data; data models; qualitative data; sharing of data. The last four terms are very important for platform context.

Table 4. Comparison top 40 Key Phrases of " all Key Phrases AND "platform" filter" from files: 'OnePetro-Words clusters by textextract platform.csv' and 'KeyPhrases_scopus-titles-abstracts-platform.tsv'

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score
data platform	145.810	digital platform	5503.811
industrial data platform	134.697	digital platforms	5481.494
software platform	60.966	social media platforms	644.251
offshore platform	44.089	mobile platform	500.571
digital oilfield workflow platform	32.671	platform design	472.958
digital oilfield solution platform	30.213	digital information platforms	345.276
one platform	21.062	digital platform capabilities	319.513
industrial data platform framo	19.898	big data platforms	260.621
offshore production platforms	19.891	digital platforms business performance	239.589
iot platform	16.528	digital services platform	239.139
integrated software platform	16.251	design platform	236.479
automated model building platform	13.261	digital construction platform	218.132
data enablement platforms	13.060	public health platforms	218.001
industrial internet of things iiot platform	12.662	different digital platforms	209.799
self-service analytics platforms	12.530	platform technology	206.227
data-driven platform	12.524	circuit design platform	188.620
platform approach	12.418	big data platform	181.924

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score
cloud-based software platform	12.369	potential digital learning platforms	167.460
single platform	10.866	spatial information management platform	164.023
new style of iiot platform	10.760	use of technology platforms	163.236
first platform	10.618	complex digital platforms	156.499
workflow orchestration platform	10.552	digital technological platforms	151.521
unique software platform	10.532	other online digital platforms	149.024
enhanced armr software platform	9.307	digital platform service providers	147.378
operations wells platforms compressors	8.775	mobile social media platform	144.016
bpm platform solutions	8.304	2d system design platform	140.586
cloud-based platform architecture	8.227	mobile health platforms	140.373
next-generation data-driven platform	7.627	analysis platform	138.336
viable allocation platform	7.540	platform development	138.133
proposed iiot platform	7.386	three design platforms	137.409
secure platform	7.319	digital platforms activities	133.919
simple platform	7.319	digital entrepreneurship process platform strategies	133.506
topside platform	6.729	technology platforms	125.021
exhibitionan integrated platform	6.676	mobile digital communications platforms	124.226
platform drives	5.922	digital platform a case of global healthcare enterprise	118.972
scalable cloud-based platform	5.437	one virtual technology platform	110.871
platform aggregates	5.351	digital platform businesses	109.458
expensive platform layer	5.331	analysis integration platform	104.773
wide market availability of bpm platform solutions	5.301	online software platform	104.494
sustainable competitive platform anders brun partner	3.043	potential implementation platform	101.953

Remark: The concept of "platform" are important for both petroleum and common domains, in petroleum domain the "IOT" term is more frequent but "analysis" is less frequent. The concept "cyber-physical systems data analytics" incorporate both "IOT" and "analysis".

Table 5. Comparison top 30 Key Phrases of " All Key Phrases AND "**image**" filter" from files: 'OnePetro-Words clusters by textextract image.tsv' and 'KeyPhrases_scopus-titles-abstracts-image.tsv'

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score
human brain processing images	3.926	digital images	3390.759
numeric curves arrays waveforms images maps	2.000	image quality	1473.262
resultant images	2.000	image processing	641.055
		image analysis	580.290
		digital image	549.698
		medical images	449.223
		information of image	333.474
		image sensor	313.926
		digital image processing architecture	268.284
		evaluation of image quality	259.154
		blue-band image information	251.534
		digital image processing technique	233.882
		high-resolution digital image data	229.776
		image quality parameters	228.113
		image processing systems	212.233
		process of image	205.014
		digital image encryption	203.724
		image acquisition	202.061
		standard digital images	175.865

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score
		digital information e.g. images	175.840
		digital radiographic image processing	175.387
		digital complex images	175.311
		validation of image analysis data	172.863
		image features	172.258
		image quality performance	167.152
		social image	157.334
		image production	154.198
		improved image quality	146.258
		image processing method	142.838
		high image quality	130.924

Remark: The concept of "image" rarely meet in 'digital transformation AND OnePetro' but 295 times in file: 'KeyPhrases_scopus-titles-abstracts-image.tsv'. There is high potential to transform experience in common digital image processing to petroleum domain

Table 6. Comparison top 30 Key Phrases of "All Key Phrases AND "remote" filter" from files: 'OnePetro-Words clusters by textextract remote.tsv' and 'KeyPhrases_scopus-titles-abstracts-remote.tsv'

OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score
remote operations management	28.485	remote digital technology	271.272
real-time remote operations	20.031	remote management	149.693
remote monitoring of heat exchanger data	15.544	remote system	98.153
remote operation	10.466	remote data exchange	76.503
remote observation of operations	9.281	remote monitoring	72.677
remote valve control	8.446	remote areas	70.899
remote collaboration	7.416	other remote technologies	70.718
remote correction of safety practices	5.993	remote control	70.347
remote surveillance alarm	5.289	remote monitoring of process parameters	66.897
remote location	4.356	wireless remote control	61.573
remote territories	2.632	remote patient monitoring	56.810
		remote monitoring of patients	51.290
		remote clinical supervision	48.432
		remote production center	39.898
		different remote means	38.960
		remote settings	36.444
		remote production units	35.835
		remote researcher access analysis	33.734
		remote provision	29.967
		six key components remote usage	29.586
		remote engineering application packages	29.507
		remote patient monitoring rpm	25.887
		remote smart oil pipeline surveillance system sopss	21.835
		remote diagnosis	21.755
		remote logins of computer networks	21.575
		remote expert	21.272
		remote sensing inc	19.768
		common olive remote	17.801
		remote hpc resources	17.095
		remote monitoring of poc instrument functionality	15.347

Remark:

- The common key phrases such as: 'remote researcher access analysis', 'remote diagnosis', 'remote expert', 'remote data exchange', 'remote monitoring of process parameters' all concern the key phrases listed in petroleum domain in particular: operation, control, management, collaboration, but for **research performance items** there are issues for petroleum domain:

'remote researcher access analysis', 'remote diagnosis', 'remote expert'.

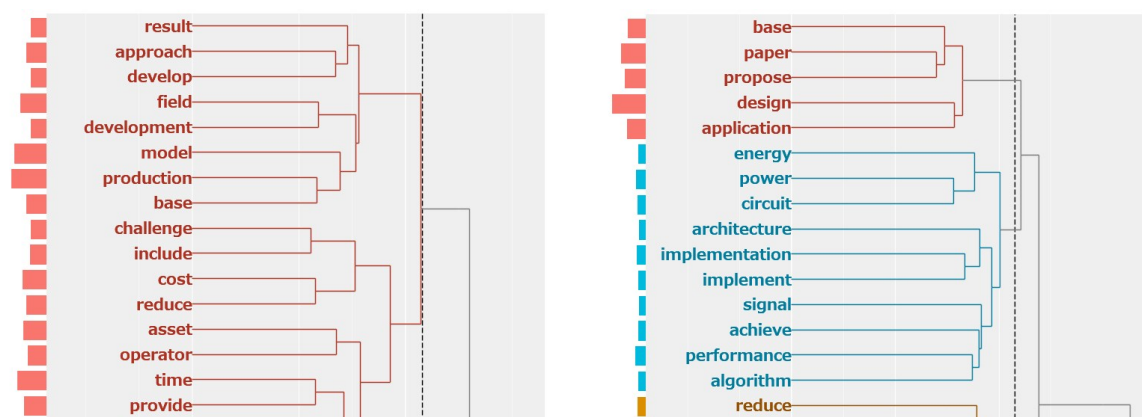
- Remote access to data for experts and researchers also has high importance for development "cyber-physical systems data analytics"

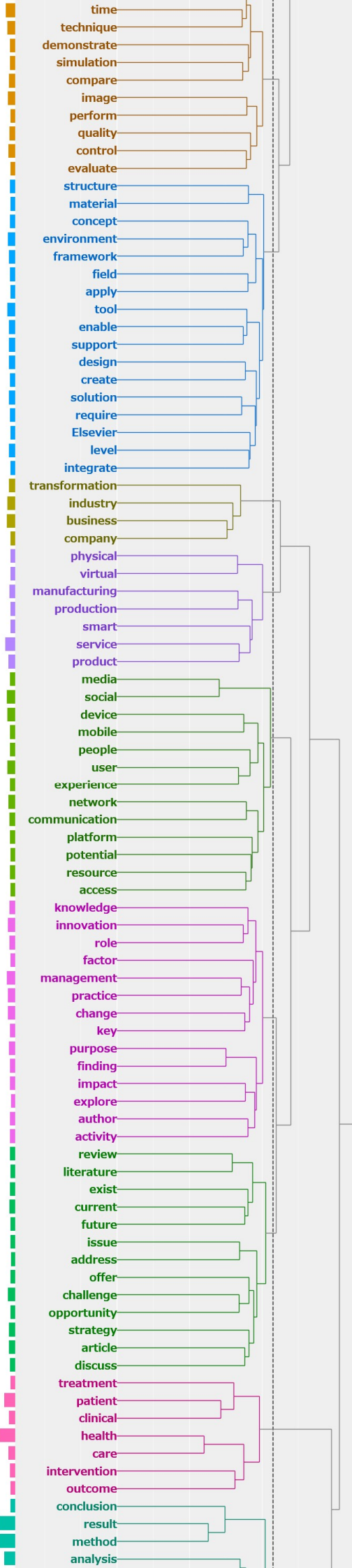
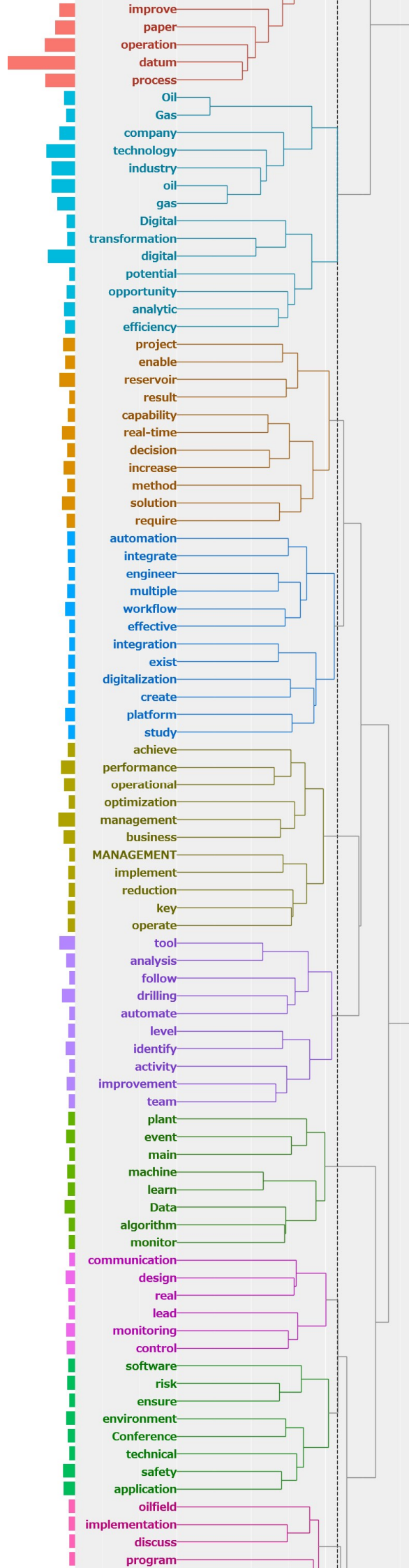
Table 7. Comparison top 30 Key Phrases of " All Key Phrases AND "sensor" filter" from files: 'OnePetro-Words clusters by textextract sensor.tsv' and 'KeyPhrases_scopus-titles-abstracts-sensor.tsv'

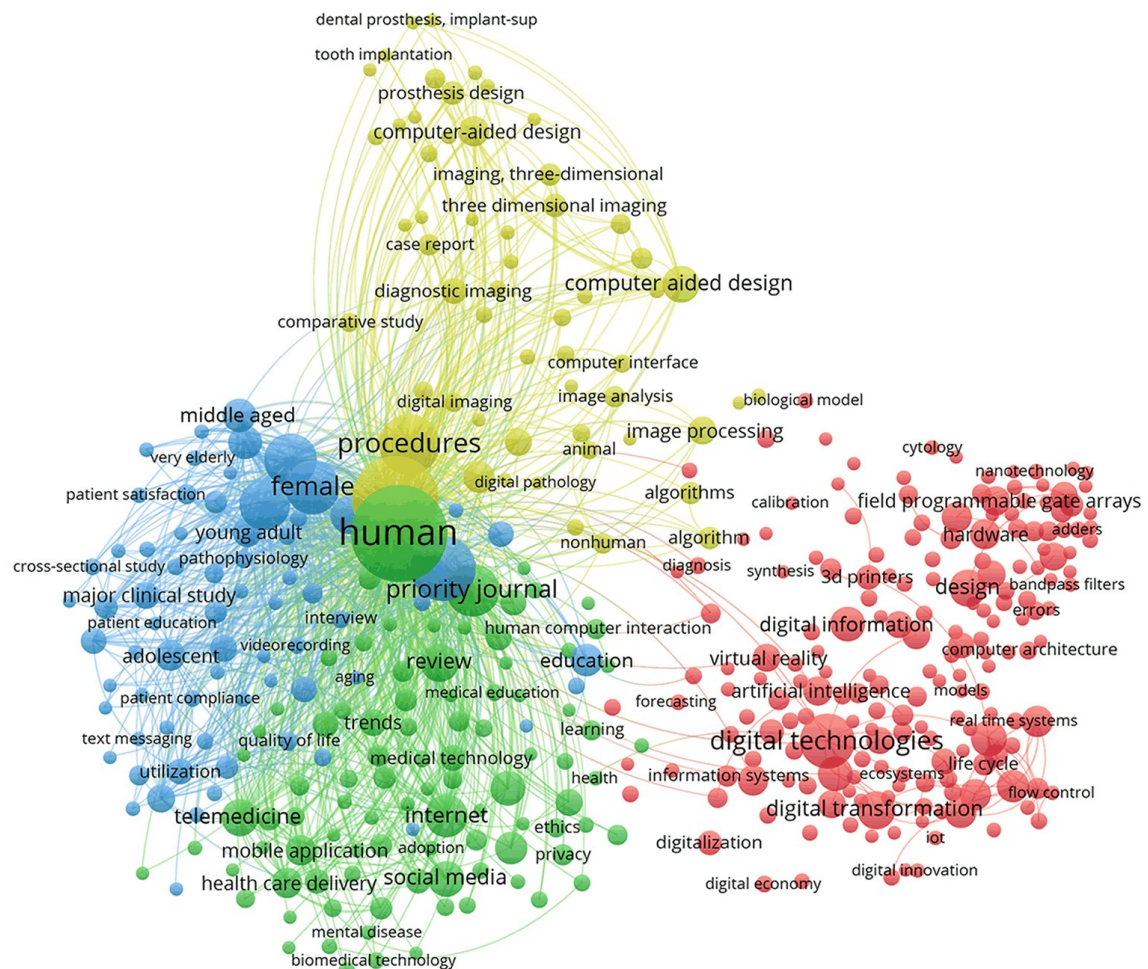
OnePetro (A) Key Phrase	Score	Scopus (B) Key Phrase	Score
sensor data	53.735	sensor data	762.203
real-time rig sensor data	32.597	new sensor	602.487
iot sensors	17.739	wireless sensor network	464.799
wireless pervasive sensor iot big data data analytics machine learning artificial intelligence etc	17.591	digital sensor	374.854
field sensor network maintenance reports	17.400	image sensor	313.926
operational sensor	17.084	wireless sensor networks	302.088
new sensor technologies	15.481	sensor technologies	264.486
analytics sensors	15.460	digital sensors	185.445
new sensors	11.961	sensory data	179.899
continuous real-time sensor	10.872	smart plant sensors	179.317
wireless sensor networks analytics software	10.733	dynamic sensor data	166.510
internet of things iot sensors	10.286	sensor data integration	157.129
number of sensors	8.365	environmental sensor data	142.149
network of sensors	8.231	sensor devices	135.290
individual time series sensor signals	8.132	real-time sensor data	134.988
human sensor	7.629	sensor network technologies	134.171
sensor installation	6.527	sensor energy	121.948
multiple sensors	6.184	wireless sensors	114.807
multiple sensors measurement	5.939	wireless sensor networks wsns	105.999
wireless sensors	4.829	sensor data fusion	102.171
conventional sensors	4.671	available digital sensors	93.139
affordable types of sensors	4.649	smart sensors	91.045
unique cost-effective sensors	3.997	digital x-ray sensors	88.850
wired temperature sensors	3.928	virtual sensors	88.814
sensor values pids erp	3.222	sensory information	80.510
affordability of sensors	2.872	health sensors smart home devices	75.445
non-intrusive sensors	2.872	current sensor	72.474
valve-control flowmeter water-cut sensor and/or pressure/temperature	2.533	digital medical world of omics data biosensors	71.389
		mobile wireless sensor networks a review	70.886
		smart manufacturing additive manufacturing internet of things flexible sensors	68.896

Remark: for petroleum domain IOT (also see table 4) is of high priority, 'valve-control' is the main device for remote control, but 'image sensor' is less in use than in common items of digital transformation

"Big" picture of key words for common and petroleum digital transformation domains as return by VOSviewer and KH Coder







er Scopus: co-occurrence, all keywords, 15607 terms, 407 meet the threshold 10 VOSviewer

Figure 2. The Network of keywords co-occurrence build by VOSviewer using:

1. OnePetro: 5617 terms, 229 meet threshold 5, selected terms 137 (60% from 229)
2. Scopus: 45414 terms, 607 meet threshold 15, selected terms 364 (60% from 607)
3. Scopus: co-occurrence, all keywords, 15607 terms, 407 meet the threshold 10

Table 8. Comparison of the results of top 50 key terms extracted from abstracts and titles of bibliographic data from Scopus using: 1. KH Coder 3, VOSviewer, author KW and WK+

- **KH Coder 3**, key words extracted from abstracts and titles of bibliographic data from Scopus
- **Scopus**, abstracts and titles: 45414 terms, 607 meet threshold 15, selected terms 364 (60% from 607)
- **Scopus**, co-occurrence: all keywords, 15607 terms, 407 meet the threshold 10

KW by KH Coder 3	KW by VOSviewer	KW and KW +
digital technologies	digital system	human
digital technology	technique	humans
digital information	performance	article
digital systems	industry	procedures
digital design	service	digital technologies
digital transformation	practice	female
digital system	device	male
digital health	patient	priority journal
digital twin	opportunity	adult
big data	structure	internet
social media	experience	controlled study
digital data	year	digital transformation
digital environment	product	review

KW by KH Coder 3	KW by VOSviewer	KW and KW +
digital media	algorithm	computer aided design
health information	state	design
case study	innovation	telemedicine
new digital technologies	operation	social media
new technologies	perspective	manufacture
digital economy	person	digital information
future research	outcome	big data
health care	architecture	technology
digital business models	group	middle aged
information technology	addition	aged
digital health technologies	rate	digital technology
digital innovation	simulation	digital twin
digital age	company	education
information systems	digital transformation	devices
digital twins	evidence	internet of things
digital designs	digital design	industry 4.0
digital pathology	image	embedded systems
new digital technology	organization	hardware
data analysis	health	mobile application
mental health	understanding	adolescent
digital solutions	case study	decision making
digital platform	function	health care delivery
simulation results	parameter	computer-aided design
digital platforms	access	signal processing
digital system design	relationship	field programmable gate arrays (fpga)
digital disruption	survey	artificial intelligence
digital devices	circuit	young adult
health system	implication	digital devices
3d printing	measurement	energy efficiency
digital health tools	computer	psychology
digital communication	insight	digital system
digital technology use	signal	innovation
digital strategies	accuracy	virtual reality
digital model	processing	image processing
digital world	property	child
power consumption	intervention	information technology
digital tools	requirement	digital storage

Remark: On my opinion, KH Coder 3 gives the most attractive results compare to VOSviewer and authors and KW+

Results:

- "Cyber-physical systems data analytics" could be an attractive emergence topic for petroleum domain based on experience in other industries. **Emphasis on analytics** for cyber-physical systems.
- The concept of "**platform**" are important for both petroleum and common domains, in petroleum domain the "**IOT**" term is more frequent but "**analysis**" is less frequent. The concept "**cyber-physical systems data analytics**" could incorporate both "IOT" and "analysis".
- KH Coder 3 gives the most attractive results by extracting key phrases from titles and abstracts compare to VOSviewer and authors and KW+

References:

KH Coder (by HIGUCHI Koichi) is a free software for **quantitative content analysis** or **text mining**. It is also utilized for **computational linguistics**. <https://khcoder.net/en/>

Van Eck, N.J., & Waltman, L. (2010). Software survey: **VOSviewer**, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.

Supplementary material on Figshare:

