

DOE Bioenergy Technologies Office (BETO) 2021 Project Peer Review

3/16/2021

Feedstock-Conversion Interface Consortium

Task 4 – Data Integration and Management

Jim Collett (PNNL) and Rachel Emerson (INL)

This presentation does not contain any proprietary, confidential, or otherwise restricted information



FCIC Task Organization





Project Overview



Objectives:

- Provide a web-enabled database for integrating, standardizing, and archiving FCIC data intended for public release, and for channeling integrated datasets to Task 8 Cross-cutting Analysis.
- Establish controlled vocabularies, data standards, and critical property specifications for biorefinery unit operations and process streams within a Quality by Design (QbD) framework.
- <u>Deploy a public web portal</u> to disseminate FCIC datasets, specifications, analyses, methods, and software to industry stakeholders who will grow the new bioeconomy.



Task 4 – Data Integration

Project Overview

- **Current limitations:**
- A lack of data standards and public, commercial-quality data servers holds back the bioeconomy by limiting dissemination of BETO-funded datasets and software tools for use by equipment designers and biorefineries.
- <u>FCIC R&D teams do not have a modern digital platform</u> to collaboratively assemble and publicly share the massive amount of data they generate, and still rely heavily on emailing spreadsheets or share drives to exchange data.

Relevance:

- This enabling task seeks to make every researcher in the FCIC more efficient and effective by providing a modern, commercial-quality platform for collaboration and data sharing across 8 National Labs.
- BETO-funded datasets, models, and QbD tools are now <u>Findable, Accessible, Interoperable, and Reusable</u> (FAIR) in accordance with DOE mandates for data modernization.
- This task directly supports 3 (of 5) major recommendations from the FCIC 2019 Peer Review report (p. 67):
 - Develop innovative ways to <u>immerse the national laboratories with industry</u>.
 - Establish <u>quality specifications</u> on feedstocks
 - Aggressively promote FCIC results and visibility.

Risks:

- <u>Sustained commitment to data standards</u> and <u>timely data uploading</u>, integration, and dissemination are essential for encouraging stakeholder confidence in the FCIC.
- Meeting <u>User Experience expectations</u> and maintaining near <u>100% uptime</u> are necessary to turn new Data Hub users into devoted fans of FCIC knowledge products and tools.







Argonne 🕰

mm

BERKELEY LAB









Pacific Northwest

Sandia National

1 – Management





Subtask	Lead(s)	Major Responsibilities
4.2	Jim Collett (PNNL) and Dave Sievers (NREL)	FCIC Data Hub Infrastructure, Tools, and Training
4.4	Rachel Emerson (INL) and Dave Sievers (NREL)	Facilitating Task 8 TEA

Risk Mitigation:

- LabKey Premium Edition Server Software on the FCIC Data Hub provides:
 - <u>enterprise-grade</u>, <u>commercial-quality web application</u> software deployed on the Amazon Web Services cloud for excellent uptime and availability.
 - <u>Same-day technical support</u> and weekly meetings with LabKey developers.
- FCIC Data Hub Content Management
 - <u>Separate "production" and "development" servers</u> are maintained by a professional IT team at PNNL to minimize risks of code deployment and software upgrades
 - <u>Datasets and code are compatible with the free, open-source Community</u> <u>Edition of LabKey Server</u> to allow legacy access to data if funding for the consortium is discontinued.

Communication strategy:

- Core Data Hub developers, data scientists, and IT staff coordinate via <u>Microsoft</u> <u>Teams</u> and <u>LabKey's built-in issue tracking system</u>.
- <u>Funding is provided for "Data Liaisons" at all FCIC NLs</u> to attend biweekly meetings and annual meeting.
- <u>Eight hours of online LabKey user and developer training</u> is made available to FCIC researchers each fiscal year.
- <u>Feedback sessions with IAB</u> and phased public roll-out during planned for FY21.



2 – Approach



Technical Approach:

- LabKey workflows for integrating bench-to-PDU scale experimental data and feedstock characterization data within "Study" data structures have been created on the Data Hub.
- Each of the FCIC's 35 Subtasks has its own Study folder wherein datasets, statistical views of the data, figures, and text may be collaboratively compiled into publications supported by curated, downloadable datasets.

Biomass Feedstock National User Facility

Biomase Material

Values Search

and a plan when mind when particles

Bioenergy Feedstock Library

GUID List Search

Sample Database

Biomass Data Tools



2 – Approach



Technical Challenges:

- R&D teams have for generations worked within "silos" within their institutions; finding consensus on interoperable data formats, definitions, and protocols to enable the success of the FCIC is an ongoing but surmountable challenge.
- <u>Data standards and definitions must originate with the researchers</u> who are experts in their field—some with decades of experience.
- Striking a balance to <u>encourage FCIC researchers to share data</u> via the Data Hub while still allowing for publishing advantage and patenting.

Technical Metrics:

- LabKey Server is an FDA-compliant database with password protected, role-based permissions and extensive activity auditing tools that describe in detail who is using the Data Hub, how often they are using it, and what data and information they are accessing.
- <u>LabKey's built-in survey tools</u> will be used to collect feedback from stakeholders during public roll-out in FY21.



3 – Impact



Impact:

- FCIC research teams across 9 National Labs now have:
 - <u>A modern, commercial-quality collaborative platform</u> to integrate, standardize, and archive datasets for advanced analytics and TEA/LCA.
 - A web portal with <u>advanced search tools for sharing</u> <u>FCIC datasets</u> with industrial stakeholders.
- The Data Hub's QbD database and searchable Subtasks and publications will provide <u>well-documented data and</u> <u>evidence</u> for:
 - <u>Material properties</u> for biorefinery feedstocks, intermediates, and products.
 - Process parameters for commercial unit operations to facilitate biorefinery scale-up.
 - <u>Data that can support market specifications</u> for feedstocks, intermediates, and products throughout the value chain.



LabKey dataset being used as training data for machine learning to dissect corn stover bale quality in real time within Subtask 2.4, Feedstock Variability At The Macro-Scale. "BFL GUID" column entries directly link to samples in the Bioenergy Feedstock Library at INL.

8

3 – Impact



Dissemination:

- We will promote the <u>FCIC</u> <u>Data Hub</u> as the "go-to" platform for datasets on feedstock variability and biorefinery unit operations via webinars, publications, and conference presentations.
- High quality, curated datasets on the Data Hub will be convincing "calling cards" for industrial partnerships that lead to commercialization and deployment of BETOfunded research.



9



4 – Progress and Outcomes



4 – Progress: Finding Knowledge and Tools on the FCIC Data Hub



- The <u>Data Hub Web Portal</u> is now online and equipped with a <u>"Data Finder" dashboard</u> for accessing "Study" folders for each FCIC Subtask. FCIC publications may be accessed via the "Publications" tab.
- Data accruing on the Data Hub are easily Findable and Accessible (per FAIR data guidelines) via a faceted search tool that allows
 users to click on metadata tags to rapidly down-select to Subtasks covering specific R&D focus areas.



<u>Subtask metadata tags</u> now include a range of 260 descriptors for analytical methods, feedstock types, product types, technology pathways, process areas, unit operations, process parameters, process stream properties, fiscal years active, and participating organizations.



Task 4 – Data Integration ¹¹

4 – Progress: LabKey Subtask Studies





Technical Approach:

- Each of the FCIC's 35 Subtasks has its own "Study" home page on the Data Hub
- Studies have links for datasets, analyzed data views, results narratives, provenance, publications, and metadata associated with the Subtask.

 Subtask Studies provide supporting data and information for the 66 "Knowledge" and "Tool" products developed thus far by the FCIC.



Task 4 – Data Integration

4 – Progress: Subtask Data Access and Export



- Harmonized, integrated datasets supporting FCIC publications may be <u>easily found, filtered, sorted, and downloaded</u> to Excel or to TSV or CSV text files.
- The integrated LANL, NREL and INL dataset shown below highlights how the Data Hub enables <u>self-service access to the past</u> work of others to supply **Interoperable** and **Reusable** FCIC data in accordance with FAIR data principles.

Sustainable Chemistry Engineering		Data Hub														C	2 🜣	💄 Jim	Collett
Impacts of Inorganic Material (Total Ash) on Surface Energy,	FCIC V Task Portals V St	tudies 🗸						Overview Datas	ets ()ata Viev	vs R	esults Na	rratives	Workb	ooks	Samples	File	s M	anage
Wettability, and Cohesion of Corn Stover Juan H. Leal [*] Extrella L. Torres, William Travis Rouse, Cameron M. Moore, Andrew D. Sutton, Amber N. Hoover, Chenlin L. Michael G. Resch, Bryon S. Donehoe, Allison E. Ray, and Troy A. Semelsberger [*]	Corn Stover N2 Surface	e Area and Comp	ositional Analys	sis D.	03.05.P	1 Bulk Flow Characteriz	zation												
Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable Overs: Eng. 2020, 8, 2661–2072 Cone Table: ACS Sustainable:	■ - ▲ - ⊖																	1 - 37 of	37 -
ANSTRACT: The impacts and varability of inorgatic material (measured as total and) on surface array, writers energy, writability and cohesion of const stover amplies from lows were examined with increasing als content, in puricially, the and component of writers energy with determined to be significant with the add component of writer energy with determined to be significant with the add component of writer energy with determined to the dispersive match the compositions of corn stover with writers mount of total langenic matter increasing energy. The work of obvious for the compositions of corn stover with uncreasing total and content from stover with uncreasing and discontinuous line patients. Both content, Waching arching, and discontinuous line patients. Both content, Waching proved discust we removing extinct incomparise material from cons stover with a horistic from 204% to 62%, accompatied by a significant reduction (from 85.5 to 42.5 nt/(m ²)) in the add component of the market reduction from 204% to 62%, accompatied by a significant reduction. Firstheling	Excel Text Seript Excel Workbook (.xisx) Maximu Excel Old Binary Workbook (.) Refreshable Web Query (.iqy) Column headers: ⁷ Caption V Export	um 1,048,576 rows and 16,384-co Kls) Maximum 65,536 rows and 2	lumns. St columns.																
FCIC publication accessible on Web	Experiment Name ©	Composition Sample ID	Sample INL GUID	Ash Class	Moisture Class ©	LANL GUID	Average N2 Surface Area M2 G	Assay NREL GUID Date	Ash wt%	TX ug/g	Al2O3 wt%	CaOF wt%w	203 K20 t% wt%	Mg O wt%	Mn O N wt% w	a20 P2 t% wt5	05 Si O2 % wt%	? Ti O2 wt%	SO3 wt%
n search San Amaliant Annual Maria Constanting (All San Amalian San Amalian San Amalian San Amaliant Am	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale11_Segment1	ca03f467-348d-e944- 94c4-9b4dff2cc3fc	LA	LM	ca03f467-348d-e944- 94c4-9b4dff2cc3fc	0.883	2019-05-23 61d60b5d-87ad-804 b9e3-59acfc2e5a62	le- 7.95	1720.0	5.49	6.04	1.79 10.6	6 3.58	0.06	0.79 2	.49 64.8	32 0.20	6 1.35
2 ματο το τ	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale11_Segment4	fa8f3731-0154-1b41- 8f89-b4b20808bc93	LA	LM	fa8f3731-0154-1b41- 8f89-b4b20808bc93	0.746	2019-05-23 61f4e57e-ece4-584 a0b4-10f86f7e698d	⁸⁻ 8.94	1750.0	6.32	5.92	1.99 8.9	4 3.27	0.08	0.99 2	.56 65	.8 0.21	3 1.31
	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale11_Segment6	f2772330-12f2-bc40- a6d0-218d1051bb4e	LA	LM	f2772330-12f2-bc40- a6d0-218d1051bb4e	0.732	2019-05-23 dcbcc9a4-525d-294 a85d-aa0c5a01853	d- 8.07	1730.0	5.57	6.09	1.71 11.0	9 3.22	0.06	0.91 2	.74 67.1	17 0.2	5 1.38
Mark Line Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark Mark	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale12_Segment2	4c383541-55dc-2441- 91e0-a79f5a6c9060	LA	LM	4c383541-55dc-2441- 91e0-a79f5a6c9060	0.842	2019-05-23 372bf554-6f1f-d742 bc74-8c85daba5cd	9.48	1630.0	6.58	5.41	2.02 9.	5 3.0	0.08	1.08 2	.37 68.1	18 0.2	9 1.26
	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale12_Segment3	cfe01cb9-8d28-f34c- 8987-970c4aafc541	LA	LM	cfe01cb9-8d28-f34c- 8987-970c4aafc541	0.667	2019-05-23 ad54c824-1ee6-564 8981-6aec8bdb566	2- 5.57	1170.0	5.07	7.52	1.61 13.2	2 4.4	0.07	0.78 3	.41 59.1	12 0.2	3 1.62
Control Control <t< td=""><td>LT Baseline Ash Speciation Analysis After 2nd Stage Grinder</td><td>LMLA_Bale12_Segment4</td><td>1c2a8dcf-1ea4-a448- 8a9b-591589f2502d</td><td>LA</td><td>LM</td><td>1c2a8dcf-1ea4-a448- 8a9b-591589f2502d</td><td>0.772</td><td>2019-05-23 4fa560ea-b90d-da4 a5ab-077afca7d535</td><td>f- 6.73</td><td>1490.0</td><td>5.29</td><td>6.9</td><td>1.77 12.9</td><td>3 3.71</td><td>0.08</td><td>0.78 3</td><td>.56 61.5</td><td>i5 0.2(</td><td>ó 1.62</td></t<>	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale12_Segment4	1c2a8dcf-1ea4-a448- 8a9b-591589f2502d	LA	LM	1c2a8dcf-1ea4-a448- 8a9b-591589f2502d	0.772	2019-05-23 4fa560ea-b90d-da4 a5ab-077afca7d535	f- 6.73	1490.0	5.29	6.9	1.77 12.9	3 3.71	0.08	0.78 3	.56 61.5	i5 0.2(ó 1.62
Marting Ma Terror Marting Mart	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale12_Segment5	cb28b1a0-8251-cf4e- 8188-caf0ba6fee5b	LA	LM	cb28b1a0-8251-cf4e- 8188-caf0ba6fee5b	0.79	2019-05-23 7a885eba-e618-6d4 a56c-ad8036844c3	1- 9.21	1690.0	6.18	5.65	1.94 9.4	6 3.02	0.07	0.91 3	.01 66.3	39 0.21	9 1.33
Data exported to Excel	LT Baseline Ash Speciation Analysis After 2nd Stage Grinder	LMLA_Bale12_Segment6	49d5d3a1-798b-3f4c- 9dbb-39733790aa59	LA	LM	49d5d3a1-798b-3f4c- 9dbb-39733790aa59	0.839	2019-05-23 b89f5cf9-0bde-044 9241-9ae552c936a	8- 8.04	1470.0	6.03	6.0	1.99 10.	9 3.16	0.08	0.84 3	.14 65.3	31 0.3	3 1.47

Task 4 – Data Integration

4 – Progress: Data Hub Analytical Tools



Data Analysis: The Task 4 Team is developing reusable data analysis and automated data processing tools using LabKey's extensive Application Programming Interfaces (APIs) for JMP/SAS, R, Python, Perl, Java, and JavaScript.

																					1 - 37 c	f37 =	
ixcel Text Script																							
Java JavaScript Perl R SAS Stable URI	O Python																						
riment Name	Composition Sample ID	Sample INL GUID	Ash Class	Moisture Class	LANL GUID	Average N2 Surface Area M2 G	Assay Date	NREL GUID	C.	Ash wt%	TX ug/g	Al2O3 wt%	Ca O wt%	Fe2O3 wt%	6 K20 wt%	Mg O wt%	Mn 0 wt%	Na20 wt% ©	P205 wt% ©	Si O2 wt%	Ti O2 wt%	S03 wt% ©	
seline Ash Speciation Analysis 2nd Stage Grinder	LMLA_Bale11_Segment1	ca03f467-348d-e944-94c4- 9b4dff2cc3fc	LA	LM	ca03f467-348d-e944-94c4- 9b4dff2cc3fc	0.883	3 2019-05-2	3 61d60b5d-87ad b9e3-59acfc2e	d-804e- 5a62	7.95	5 1720.0	5.49	9 6.0	4 1.	79 10.0	6 3.5	8 0.0	5 0.79	2.49	64.82	2 0.20	5 1.35	
seline Ash Speciation Analysis 2nd Stage Grinder	LMLA_Bale11_Segment4	fa8f3731-0154-1b41-8f89- b4b20808bc93	LA	LM	fa8f3731-0154-1b41-8f89- b4b20808bc93	0.746	6 2019-05-2	3 61f4e57e-ece4- 10f86f7e698d	-5848-a0b4	4- 8.94	1750.0	6.3	2 5.9	2 1.	99 8.9	4 3.2	.7 0.0	8 0.99	2.56	65.8	8 0.28	3 1.31	
seline Ash Speciation Analysis 2nd Stage Grinder	LMLA_Bale11_Segment6	f2772330-12f2-bc40-a6d0- 218d1051bb4e	LA	LM	f2772330-12f2-bc40-a6d0- 218d1051bb4e	0.73	2 2019-05-2	3 dcbcc9a4-525d a85d-aa0c5a01	N2 S	uface	Area \	/s. Cor	mpos	ition	© T.03.0	5.P1 Bulk	Flow Cha	racterizati	on				
seline Ash Speciation Analysis		4c383541-55dc-2441-91e0	-l		4c383541-55dc-2441-91e0	0.842	2 2019-05-2	372bf554-6f1f-0 8c85daba5cdf	Report	Source	Help												
aseline As	ins 🗸	Describer Distants Data	Views	unite Marrathan	arkhonin Samular Elur Hannen	0.667	7 2019-05-2	ad54c824-1ee6 8981-6aec8bdb	Scri	ipt Source													
useline As 2nd Stag N2 Suface Area vs. Com	position Character	erization				0.772	2 2019-05-2	3 4fa560ea-b90d 077afca7d535	1 2 3	library	(Rlabk	ey); vfile}											
aseline As Report Source Help r 2nd Stag SVG output						0.79	9 2019-05-2	3 7a885eba-e618 a56c-ad803684	5 6 7	fol sch	data seUrl=" derPat	https:/ h="/FCI	/bioen C/FCIC	ergy.la Task W	bworks. ork/T.0	org/lab) 3.00 Mat	cey", cerials	Handling	T.03.0	5.P1 Bu	lk Flow	Charact	eriza
aseline As		•				0.839	9 2019-05-2	3 b89f5cf9-0bde- 9ae552c936a7	8 9 10	que vie col	eryName wName= LSelect	="Surfa "" ="Subje	ce_Com	positio sh Clas	n_Join_ s,Moist	GUID_Mat	ch", ss,Sampl	e Descrij	otion,A	verage	N2 Surf	ace	
aseline As	000	0	•		0 0	0.883	3 2019-05-2	3 68f92795-47e7 3bfd3106c9b5	11 12 13	Area_m2 col cor col	g,Ash Filter tainer NameOp	percen =makeFi Filter= t="fiel	t,Comp lter(c NULL, dname"	osītion ("Avera);	_Sample ge_N2_S	ID,SO3 arface 1	wt%", Area_m2_	g", "LES:	S_THAN"	, "2"))	,		R
seline As 2nd Stag	000	P				0.77	5 2019-05-2	3 4d8a2784-f083 96a3b55e94ca	- 14 15 16	#write.	table(labkey.	data)					(501)	~~~ ~	odo	visih	lo to i	1001
aseline As			00	0	00	1.00	7 2019-05-2	3 7935cc6d-dbe3 9751-98e6f706	17 18 19 20	<pre>library # \${svg svg("ou</pre>	(ggplo jout:ou itput.s	t2); tput.sv vg", wi	g} dth= 1	8, heig	ht=6);			(500	00 0	ouc	VISIO		001
iseline As ⁸ 7 2nd Stag ¹⁰⁰⁻					0	0.900	6 2019-05-2	3 5d171032-9598 a1b5-e55964b5	20 21	#ggplot geom_te	(data ext(siz	= labke e = 3);	y.data	, aes(x	= (),	y = (),	label =	(Moistu	re_Clas	s)) + g	eom_poi	nt() +	
seline As 2nd Stag	۲					0.800	6 2019-05-2	3 5a387168-e756 bc23-d447c614	23	ggplot Moistur geom_sr	data = ce_Class nooth(m	labkey s)) + ethod=1	.data, geom_p m, col	aes(x pint(co pr=' <mark>#2C</mark>	= (`Ave lor= 'd 3E50')	rage_N2 odgerbli + xlab()	Surface e2', si oquote("	Area_m2 ze = 8) Average 1	g`), y geom_ itroge	= (`SC text(si n Surfa)3 wt%`) .ze = 3, .ce Area	<pre>, label color=' m2/g"))</pre>	white
iseline As		e				1.052	2 2019-05-2	3 ab41b9d9-449b	25	10.50	720.0	7.4	1); 1 4.1	o	51 4.0	io 2.0	0.0	0.57	0.77	74.00	0.44	+ 0.7	

4 – Progress: Quality by Design Data Integration

Mapping QbD Properties onto Technology Pathways

- Process Stream Critical Properties
- Unit Operation Critical Process Parameters
- <u>Collaborative QbD property</u> <u>assignment</u> supports:
 - self-organization of metadata into <u>ontologies</u> within the FCIC community.
 - consensus for defining <u>controlled vocabularies and</u> <u>units of measure</u> for use across the growing bioeconomy.



Tools

4 – Progress: Quality by Design Data Integration



- A <u>Process Stream Critical Properties</u> data table was created for managing the assignment of QbD critical material properties to feedstocks, intermediates, and products within the High- and Low-Temperature Conversion pathways.
- A related <u>Unit Operation Critical Process Parameters</u> data table was created for assignment of critical process parameters to unit operations within the High- and Low-Temperature Conversion technology pathways.
- Subtask teams and the FCIC PI have reached consensus on a first batch of 164 input material attributes, 85 output quality attributes, and 79 process parameters; these data are now being harmonized and uploaded to the database.

m +	🙇 - + - 🗇 🕹 -	⊖										1 - 12 of 12 -
defa	This grid view has been modified.	Revert Edit Save										
(<u>× T N</u>	aterial Property IS ONE OF (MOISTU)	K ▼ Unit Op IS ONE OF (DRAG_ Unit Op ▼ ©	CHAIN_CON) Intermediate Stream • 0	Clear All Intermediate State	Material Property T Ö	Units ©	CMA Upper Limit ©	CMA Lower	CQA Upper Limit ©	CQA Lower	Downstream Unit Op 🗢	Critical Downstream Unit Op
0	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	009	Chips	ASH_CONTENT_01	%	30.0	4.0	30.0	4.0	FRACTIONATION_01	COMMINUTION_01
0	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	009	Chips	PARTICLE_SIZE_DISTRIBUTION_01	mm	50.0	6.0	50.0	6.0	FRACTIONATION_01	COMMINUTION_01
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	009	Chips	MOISTURE_CONTENT_01	%	35.0	10.0	35.0	10.0	FRACTIONATION_01	COMMINUTION_01
	HIGH_TEMPERATURE_CONVERSION_02	FRACTIONATION_01	010	Chips	PARTICLE_SIZE_DISTRIBUTION_01	mm					DRAG_CHAIN_CONVEYOR_01	
0	HIGH_TEMPERATURE_CONVERSION_02	FRACTIONATION_01	010	Chips	ASH_CONTENT_01	%					DRAG_CHAIN_CONVEYOR_01	
0	HIGH_TEMPERATURE_CONVERSION_02	FRACTIONATION_01	010	Chips	MOISTURE_CONTENT_01	%					DRAG_CHAIN_CONVEYOR_01	
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	011	Chips	PARTICLE_SIZE_DISTRIBUTION_01	mm	50.0	6.0)	1.0	COMMINUTION_02	PYROLYZER_FEED_SCREW_01
Ċ.	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	011	Chips	ASH_CONTENT_01	%	30.0	4.0)	0.0	COMMINUTION_02	PYROLYZER_FEED_SCREW_01
0	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	011	Chips	MOISTURE_CONTENT_01	%	35.0	10.0)	3.0	COMMINUTION_02	PYROLYZER_FEED_SCREW_01
0	HIGH_TEMPERATURE_CONVERSION_02	COMMINUTION_02	012	Particles	PARTICLE_SIZE_DISTRIBUTION_01	mm					SCREW_CONVEYOR_1_01	
	HIGH_TEMPERATURE_CONVERSION_02	COMMINUTION_02	012	Particles	ASH_CONTENT_01	%					SCREW_CONVEYOR_1_01	
Ċ.	HIGH_TEMPERATURE_CONVERSION_02	COMMINUTION_02	012	Particles	MOISTURE_CONTENT_01	%					SCREW_CONVEYOR_1_01	



4 – Progress: Access to Data for QbD Critical Property Assignment



The QbD tables support links to experimental data that may be considered in the assignment of sets of critical
properties to process intermediate streams or unit operations.

	M (*)	+ +									1 - 15 of 15 📼
× • 1	Init Op IS ONE	Lists / Material Quality Classes	RO-TAP Clear All								
	FY Active	View List Item CFCIC	Area 🔍 Unit Op 🝸	Equipment Type Control	Material Property *	Parent Assay T	CMA Upper C Limit Limit	MA Lower Ir imit © N	ntermediate Stream	CQA Upper CQA Lower	Downstream Unit Op
		Meterial or Quality Attribute: ASH CONTENT 01	ESSING ROTARY_DRYER_01	Dryer	ASH_CONTENT_01	Ash Content 750C ASTM D3174-12	30.0	4.0 0	08	30.0	4.0 DRAG_CHAIN_CONVEYOR_0
2		Material of Quality Attribute. ASH_CONTENT_01	ESSING ROTARY_DRYER_01	Dryer	MOISTURE_CONTENT_01	Moisture (%, 1050)	60.0	10.0 0	08	35.0	0.0 DRAG_CHAIN_CONVEYOR_0
0		Data Type: Scalar	ESSING ROTARY_DRYER_01	Dryer	PARTICLE_SIZE_DISTRIBUTION_01	Particle Size Distribution- Forage Separator	50.0	60.0	08	50.0	6.0 DRAG_CHAIN_CONVEYOR_0
2	18,19	Units: %	ESSING COMMINUTION_01	Hammer Mill	ASH_CONTENT_01	Ash Content 750C ASTM D3174-12	30.0	4.0 0)10a	3.0	0.0 SCREW_CONVEYOR_1_01
0	18,19	Parent Assay: Ash Percent 750C	ESSING COMMINUTION_01	Hammer Mill	MOISTURE_CONTENT_01	Moisture (% 05C)	35.0	10.0 0)10a	15.0	3.0 SCREW_CONVEYOR_1_01
0	18,19	Description:	ESSING COMMINUTION_01	Hammer Mill	PARTICLE_SIZE_DISTRIBUTION_01	Particle Size Distribution- Forage Separator	50.0	6.0 0	10a	6.0	1.0 SCREW_CONVEYOR_1_01
-	20,21	1,5 HIGH_TEMPERATURE_CONVERSION_01 PREPR	OCESSING COMMINUTION_02	Knife Mill	ASH_CONTENT_01	Ash Conten, 750C ASTM D3174-12		0	110b		SCREW_CONVEYOR_1_01
	20,21	1,5 HIGH_TEMPERATURE_CONVERSION_01 PREPR	OCESSING COMMINUTION_02	Knife Mill	MOISTURE_CONTENT_01	Moisture %, 105C)	Start Page / Assay Mar	nagement			SCREW_CONVEYOR_1_01
Assay L	st / Ash (%, 750	0C) Batches / Ash (%, 750C) Runs			PARTICLE_SIZE_DISTRIBUTION_01	Particle Size Distribution- Forage Separator	Asn Percentag	e 750C DFOC		Edit New Manag	SCREW_CONVEYOR_1_01
Ash (%, 750C) R	Results Braic			ASH_CONTENT_01	Ash Content 750C ASTM D3174-12					METERING_BIN_01
Dry-basi	s %ash measure	ed from TGA based on ASTM D 3172-07.			MOISTURE_CONTENT_01	Moisture (%, 105C)	ASDY Do	signation: D3174 - 12			METERING_BIN_01
MANAGE	ASSAY DESIGN F	VIEW BATCHES > VIEW RUNS > VIEW RESULTS > VIEW COPY-TO-STUD	Y HISTORY > VIEW EXCLUDED DATA > eplaced Filter - Exclude	1 - 7 of 7 📼	PARTICLE_SIZE_DISTRIBUTION_0	Particle Size Distribution- Forage Separator	St	andard Test Method	d for	tern 0-11	METERING_BIN_01
	efault This gr	rid view has been modified. Revert Edit Save			ASH_CONTENT_01	Ash Content 750C ASTM D3174-12	This state	estandard to issued under the flared desi- tud adoption or, in the case of territori, receipt epsilon (c) indicates an editorial	Sample of Coal and Coke ignation D3174, the monther immediately following the a , the year of last revision. A monther in parenthese indica I change since the last revision or reapproval.	designation indicates the year of nes the year of Last reapportal A	SCREW_CONVEYOR_2_01
×	Run = 685				MOISTURE_CONTENT_01	Moisture (%, 105C)	The 1. Scope 1.1 This test m	number has been approved for use by	agencies of the U.S. Department of Definite. D3172 Practice for P D3173 Test Method 1	roximate Analysis of Coal and Coke for Moisture in the Analysis Sample of	SCREW_CONVEYOR_2_01
	Experiment Na HT Baseline A	ame Sample ID BFL GUID I analysis Dry2-HAHM-D6 04a76489-db6b-da43-a60f-912accd96bef 2	Date Ash Percent 750C 2018-07-10 00:00 1.1874	Protocol Wiki Page	PARTICLE_SIZE_DISTRIBUTION_01	Particle Size Distribution- Forage Separator	ganic residue as a prepared in accorr The results obtain analysis, Practice	ash in the analysis sample of a dance with Practice D2013 or ned can be applied as the ash in D3172, and in the ultimate an	coal or coke as Practice D346. n the proximate ualysis, Practice	Itimate Analysis of Coal and Coke Calculating: Coal and Coke Analyses of to Different Bases	SCREW_CONVEYOR_2_01
	HT Baseline A HT Baseline A	nalysis Dry2-LAHM-D19 5c29330f-03b5-d447-80ae-fc47feb0c504 2 nalysis Dry2-LAHM-D3 f200d4c5-9afb-d94c-9d85-3c90dab93c41 2	2018-07-10 00:00 0.5096 2018-07-10 00:00 0.3731	Ash Percentage 750C			D3176. For the reference is made Test Method D633 for trace element to mode	determination of the consti e to Test Methods D3682, D43 857 should be used to prepare analysis. See Terminology D	auents in ash, 326 and D6349, ash to be used D121 for defini- D121 for defini-	tees from Coal Utilization Processes for Major and Minor Elements in Coal X-Ray Fluorescence for Total Sulfur in Coal and Coke	
0	HT Baseline A	nalysis Stg2-HALM-C4 c6309fbe-6cdc-1e45-8ac7-b943b5d7060a 2	2018-07-10 00:00 4.1305	Ash Percentage 750C			1.2 The values	s stated in SI units are to b	be regarded as Combustion Resid Furnace Combustion	lues Using a High-Temperature Tube on Method with Infrared Absorption	
	HT Baseline A	nalysis Stg2-HALM-C7 f5f70c25-6f66-1d48-aed9-1dc4195a9206 2	2018-07-10 00:00 2.8119	Ash Percentage 750C							
	HT Baseline A	nalysis Stg2-LALM-C4 550e0857-f909-d147-9a76-a6fee1f3315a 2	2018-07-10 00:00 0.7436	Ash Percentage 750C							A 7
	HT Baseline A	nalysis Stg2-LALM-C6 7599e95e-a198-8b48-af95-deb1b4021315 2	2018-07-10 00:00 0.6884	Ash Percentage 750C		Data show	vn are not actu	ual and are	provided for la	yout demonstration	n only

4 – Progress: Experimental Data Integration



- Screenshot shows part of a dataset containing 216 records with 224 data columns integrated via a simple SQL query from 3 uploaded data files containing:
 - Feedstock property data from INL
 - Bio-oil property data and pyrolyzer run metadata from NREL
 - Fuel property data and hydrotreater run metadata from PNNL
- Stable, reusable SQL code and user access to source files promotes trust in large datasets on the Data Hub for data mining analysis and machine learning.
- Performing such integrations manually in Excel is timeconsuming and error-prone.

	Bioene	rgy Data	Hub																										~ *	-		
FCIC 🗸 T	'ask Portals 🗸	Studies	×																							Overvi	w	Subject	Data Vi	IWS	Mana	80
FP and H1	T Joined D	ata 1 -	Sele	ected F	ields	🗅 Pyrolysis Bio-	Oil and Hydrotrea	ater Data I	ntegratio	n																						
· · ·	±- ⊖																															
PR 2021)																															
× T Py Fee	d Type IS ONE C	F ('air cla	ssified																													
INL Feedstock GUID	Py Feed Type T	HT Feed Type ©	Py FP Run Or Oil	HT Feed Provider Run Number	PNNL HT Run Number	Py Run Date 🗢	HT Run Date 🗢	Py Pyrolyzer ID	HT Provider Reactor Type	HT HT Reactor	Py FP Total Liquids Wt Pct	HT Upgraded Oil Yield g/g	Py Char Yield Pct	Py Gas Yield Pct	HT Gasoline Fraction Wt Pct	HT Diesel Fraction Wt Pc	HT Jet Fuel Fraction Wt Pc	HT Heavies Fraction Wt Pc	Py Carbon Wt Pct	HT C Wt Pc ©	Py Hydrogen Wt Pct ©	HT H Wt Pc	Py HT Nitrogen N Wt Pct Wt Pc	Py Oxygen Wt Pct	HT O Wt Pc ©	Py S Wt Pct	HT S Wt Pc	Py Water Wt Pct	HT HT Viscosity Ga C P 40C Yie g/9	HT s Carl Id Bali g g/g	rbon M lance f	IT Mass Jalance J/g
032b4e7c- 2d86-434b- 9aec- 3fa7980e2e1e	Lobiolly Pine (whole pine)	whole pine (LWP)	6116- 037	6116- 037	62006-3	1776-01-02 15:00	2015-05-19 00:00	2FBR	2FBR	PDLE	65.7	0.462	13.4	15.8	38.7412	43.8466	28.3682	17.4123	46.1	88.09	7.5	12.6	0.07 0.0	46.2	1.03	51.1159	0.0	20.5	1.056 0.1	63 0.	.8723	0.936
17113d89- e706-ae4d- 9548- b0950fac5ae9	Whole Pine (WP)	Whole Pine	5858- 021	5858- 021	61573- 43	1776-01-02 15:00	2014-06-19 00:00	2FBR	2FBR	PDLE	62.9	0.501	15.0	18.9	42.0	44.0	11.0	14.0	45.3	87.52	8.1	12.67	0.09 0.0	46.5	1.03	41.0	0.0	22.5	1.0545 0.1	41 0.4	.8763	0.99
3138a563- ace0-eb46- 8658- 0a209687545e	Tulip Poplar (TP)	Tulip Poplar Repeat	5858- 074	5858- 074	61573- 51	1776-01-02 15:00	2014-07-31 00:00	2FBR	2FBR	PDLE	70.6	0.401	7.6	13.5	40.0	46.0	12.0	14.0	44.8	86.05	7.7	11.99	0.08 0.0	47.4	0.69	48.0	0.0	18.8	1.2003 0.1	91 0.1	.8368	0.95
427efecf-f5a2- 974c-b19a- 5918a53ec5d2	Hybrid Poplar (HP)	Hybrid Poplar	5858- 020	5858- 020	61573- 45	1776-01-02 15:00	2014-06-26 00:00	2FBR	2FBR	PDLE	64.9	0.46	8.9	19.2	45.0	42.0	11.0	13.0	45.9	86.76	8.1	12.42	0.08 0.0	45.9	1.01	56.0	0.0	20.7	0.9677 0.1	72 0.1	.8821	1.00
88e73d0e- 9e8d-a842- 8c41- 4e8125a90dc7	Lobiolly Pine (CP)	Clean Pine	5858- 023	5858- 023	61573- 42	1776-01-02 15:00	2014-06-19 00:00	2FBR	2FBR	PDLE	65.2	0.515	12.0	17.9	39.0	44.0	12.0	17.0	45.0	87.15	7.8	12.516	0.08 0.0	47.1	1.082	37.0	0.0	21.1	1.2186 0.1	52 0.'	.9275	1.01
972be12c- a954-624f- 955f- 735c83a95f2d	Lobiolly Pine (CP)	Loblolly pine	6408- 001	6408- 001	62006- 42	1776-01-02 15:00	2016-08-02 00:00	2FBR	2FBR	PDLE	70.7	0.443	10.5	13.6	43.8531	29.0193	9.2955	88.6177	46.59	88.24	7.08	12.58	0.04 0.0	46.3	0.728	49.0	0.007	18.1	1.1916 0	22 C	0.911	0.99
977176f3- 92e4-cb40- a2d2- 0b217c3aa061	Lobiolly Pine (clean pine)	clean pine (LPC)	6116- 020	6116- 020	61573- 72	1776-01-02 15:00	2015-03-02 00:00	2FBR	2FBR	PDLE	63.1	0.481	12.3	17.7	41.4432	45.0307	29.7181	13.5261	47.0	87.82	7.5	12.82	0.06 0.0	44.9	1.09	51.2521	0.02	21.7	0.945 0.1	91 C	0.914	0.99
cec77c74- 4ce7-734a- bd28- 9d14ec81530c	Forest Residue (FR)	Forest Residue	6207- 102	6207- 102	62005- 40	1776-01-02 15:00	2016-06-07 00:00	2FBR	2FBR	PDLE	62.0	0.466	19.1	16.5	46.5527	42.9406	29.111	10.5067	40.19	88.63	7.87	12.38	0.19 0.0	51.75	1.044	158.0	0.016	16.4	1.2532 0.1	56 1	1.045	1.
df5ebbcc- b8de-ba48- 8cc7- c245af6b5b35	air classified Forest Residue (acFR)	air cleaned FR	6408- 080	6408- 080	62006- 48	1776-01-02 15:00	2016-09-16 00:00	2FBR	2FBR	PDLE	67.5	0.491	13.1	14.8	42.9938	28.2799	12.4186	90.4751	48.48	87.97	7.48	12.38	0.03 0.0	44.0	1.0	87.0	0.0079	17.3	1.3633 0.1	72 (0.912	1.
e717596a- c33b-a745- ae21-	Pinion_Juniper	pinion juniper (PJ)	6116- 011	6116- 011	61573- 74	1776-01-02 15:00	2015-03-10 00:00	2FBR	2FBR	PDLE	57.9	0.176	15.3	17.5	60.5703	37.3962	37.0407	2.0336	50.3	86.86	8.6	14.03	0.49 0.0	40.3	1.13	245.7353	0.0	27.6	0.495 0.1	78 0.	.3784	0.897



Creation of LabKey Datasets directly from Bioenergy Feedstock Library (BFL) exports



Tools Conversion

Description

LabKey's Python API was used to develop an automated process for creating Study datasets for batches of sample provenance and analytical data exported from the Bioenergy Feedstock Library (BFL).

Value of new tool

The tool reduces the need to manually create redundant data structures between the two systems and represents an additional method to establish interoperability between the BFL and LabKey.

Potential Customers & Outreach Plan

INL researchers can more easily import samples and data directly into LabKey that are being tracked for the FCIC in the BFL database system.





Task 4 – Data Integration

4 – Progress: Milestones



	• FY19 Q2 [•] Initial funding received for this new FY19-21 Task
Task Status Snapshot	 FY19 Q3: FCIC LabKey Data Hub construction begins on AWS cloud. LabKey also installed locally
 LabKey Studies created for 35 Subtasks 	 behind firewalls at PNNL, INNL, and NREL for internal data integration and workflow development. FY19 Q4: Data from multiple NL teams integrated on Data Hub within 6 "Pioneer" Subtask Studies; interoperability with Bioenergy Feedstock Library database at INL established.
 30 publications indexed 	 FY20 Q1: A Quality by Design roadmap for FCIC is developed with input from PI and Task leads. FY20 Q2: Preprocessing Area data integration and analysis workflows constructed within LabKey
 QbD database property 	Assay and Study data structures on the Data Hub.
 assignment underway: 164 material attributes 	 FY20 Q3: Pretreatment Area data integration and analysis workflows constructed within LabKey Assay and Study data structures on the Data Hub.
85 quality attributes79 process parameters	 FY20 Q4: Quality by Design database framework deployed on Data Hub for assigning critical properties to unit operations and intermediate streams within biorefineries.
 119 Data Hub user accounts 	 FY21 Q1: Data Hub Web Portal ready for phased roll-out with FY18-FY20 reports, curated datasets, and publications organized into 35 Subtask Study folders accessible to a limited group of industry and academic stakeholders for Beta testing and feedback.
 8 hours of user and 	• FY21 Q2: Support Task 8 Case Study 1 by uploading and transforming data generated by FCIC
developer online training	experimental tasks into datasets with harmonized units and variable names to facilitate analysis.
provided to 60 FCIC	• FY21 Q3: Support Task 8 Case Study 2 by uploading and transforming data generated by FCIC
researchers during FY20	experimental tasks into datasets with harmonized units and variable names to facilitate analysis.
roboaronoro daning r 120	• FY21 Q4: Data Hub Web Portal fully open to public and updated with FY21 reports, curated datasets,
	and publications.



20





Management:

- LabKey Data Hub on AWS cloud is the central collaboration platform for integrating, standardizing, and publicly releasing datasets from the FCIC.
- Development/Admin team at PNNL, INL, and NREL; funding provided for Data Liaisons at other FCIC NLs to attend twice-monthly meetings, FCIC annual meeting, and 8 eight hours of online LabKey training each year.
 Technical Approach:
- Each FCIC Subtask has its own LabKey Study folder into which experimental and feedstock characterization data may be uploaded and harmonized to support FAIR data sharing, reuse and integrated analysis across the FCIC.
- The Data Hub Web Portal provides faceted search tools to help users quickly down-select to Subtasks covering feedstocks, intermediates, products, unit operations, QbD properties, and publications of interest.

Impact: The QbD database will provide industrial stakeholders with hundreds of well-documented specifications for material properties of biorefinery feedstocks, intermediates, products, and process parameters for unit operations to help to define operating envelopes that ensure reliable biorefinery scale-up and performance.

Progress: Data Hub Web Portal is ready for welcoming 1st round of "beta" users such as the FCIC Industrial Advisory Board, DFO partners, and BETO-supported R&D teams outside of the FCIC. It will be continually updated with data and knowledge products to support new FCIC journal publications, Task 8 Case Studies, and QbD critical property assessments.



21

Quad Chart Overview- FCIC, Task 4



Timeline

10/1/2018 - 9/30/2021

	FY20	Active Project
DOE Funding	\$559K	FY19- \$795K FY20- \$559K <u>FY21- \$489K</u> Total- \$1,843K

Project Partners (N/A)

Barriers addressed

Ct-N: Multiscale Computational Framework toward Accelerating Technology Development

Ft-E: Feedstock Quality: Monitoring and Impact on Preprocessing and Conversion Performance

ADO-A: Process Integration

Project Goal

Provide a collaborative computational environment for hypothesis development, experimental and modeling workflow management, integration of datasets and metadata, and deliverables sharing between FCIC subtasks within a uniform Quality by Design framework, as well as a portal for public access to FCIC results, data, and software.

End of Project Milestone

Data Hub Web Portal goes live and opens to industry stakeholders and public with FY18-FY21 reports, curated datasets, and publications.

Funding Mechanism (N/A)





Thank you!

Task 4 Team: David Sievers (NREL), Robert Kinoshita (INL), Shaun O'Leary (PNNL), Matt MacDuff (PNNL), Prerna Prateek (INL), Lorenzo Vega-Montoto (INL), Stan Martin (ORNL), Bruce Wilson (ORNL), Alan Chappell (PNNL)

Data and Design Liaisons: Oslo Jacobson (LBNL), Nick Dylla (ANL), Troy Semelsberger (LANL), Ken Sale (SNL), Steve Phillips (PNNL), Erin Webb (ORNL), Dave Thompson (INL)

LabKey Software Support and Developer Team: Steve Hanson, Hannah Brakke, Adam Rauch, Chet Chopra

energy.gov/fcic





Additional Slides



Task 4: FCIC Data Development Workflow





LabKey study standardization



Standardization of LabKey Study metadata

	■ • • • •	4 - A													1-	6 of 6 →
	Task 5 Studies															
	Study 🔷 💿	Start Date	Investigator	Biomass Type	Desc	cription			0	Laboratory	Task *	Subtasi	Unit Operation	Hypothesis ©	Publication DOI ©	Report ©
0	FY 18 HT Baseline Study	2019-06-05 00:00	INL	Lobiolly Pine	This the s temp	study contains all of the samples collected from perature baseline runs.	e analytical of the preproce	lata and INL PDU ssing operations	processing data for of the high	INL	2, 5		INL PDU Grinders INL PDU Dryers		10.1021/acssuschemeng.9b06718	8
0	FY18 LT Baseline Study	2019-05-23 00:00	INL	Corn Stover	This equip the lo	study contains all of the private study contains all of the private state of the same owner at the state of t	ne analytical o oples collecte ne runs.	lata and INL PDU d from the prepro	processing cessing operations of	INL	2,5		Grinders			
	FY19 CQA HT Pathway Identification Study	2020-01-28 00:00	Vicki Thompson	Lobiolly Pine	In thi pathy to ga prepr dowr runs, exper proce biom provi (CM/	is milestone, critical qu ways are identified. Th ther information on bi storeasing operations t instream conversion op , FY19 experimental da rrimental data collecte esses and from a sear ass variation on both ide context for each of As) that impact those (ality attribute ese CQAs are produce qu erations. This ta, historical f for both hig ch of publish- nigh and low the CQAs ide CQAs are pres	es (CQAs) for high the final results is all attributes and allty attributes wh is milestone draws INL BFNUF data of the BFNUF data of and low tempera- d data that desc temperature conv intified, the critica sented as well as	-temperature rom a yearls ng effort how they in Jeract with ich are pailsed to on FY18/baseline ollected/NREL ature conversion tibes the effects of ersion processes. To I majerial attributes pritcal process	INL	5	5.2	HT conversion pathways	CQAs can be identified for various HT conversion pathways through consensus of conclusion drawn		FY19 Q4 Task 5
nve	estigator ©	Biomass Type 💿	Descrip	tion	0	Laboratory	Task ^ ©	Subtask	Unit Operation	Hypot	nesis	Pu	blicatio	n DOI	© Re	eport C

Description

Creation of FCIC experimental study specific metadata parameters collected about each Study

Value of new tool

Can potentially facilitate higher level data queries and harmonization.

Potential Customers & Outreach Plan

FCIC researchers and leadership. This same concept will be applied to the Public Facing Web Portal (FY21) in order to standardize various studies and research hosted through this database.

Kolo National Laboratory



Task 4 – Data Integration ²⁶

Database Standardization & Streamlining



EEDSTOCK-CONVERSION INTERFACE CONSORTIUM

LabKey Core Data Infrastructure Improvements

ID T sampl

3 07

		Sample ID	Parent sample ID	Date C	Analyte	Parameter	Value Unit C
analvte.	0	37795		2019-12-26 00:00	acetic_acid		3.07 g/L
ata.value) AS value		37798		2019-12-26 00:00	acetic_acid		3.04 g/L
Data BY Data.sample id.		37801		2019-12-26 00:00	acetic_acid		2.98 g/L
parent_id,		37804		2019-12-26 00:00	acetic_acid		2.95 g/L
value BY analyte							
value BY analyte		Pi	voted Resu	ılts (filtere	d for o	ne sar	nple ID)

Description

Analytical data stored in efficient "long" tabular format consistent with industrial database standards. A specimen can have multiple entries that comprehensively documents each measurement done.

Value of new tool

icose total xvlose

2.44

0.95

10.75

This database format offers future flexibility and documents other data in addition to simple analyte information such as type of measurement, units, and sample lineage.

Potential Customers & Outreach Plan

FCIC researchers benefit from comprehensive analysis documentation. This same database design can be applied within other projects at national labs that use LabKey.



Subtask 4.4: Facilitating Task 8 TEA

Rachel Emerson (INL) and Dave Sievers (NREL)





4 – Progress: Assay Data Integration



Getting Data into LabKey

- Assay workflows and data structures for uploading experimental data were developed for 6 "Pioneer" Subtasks.
- Stable data table structures and harmonized data types facilitated rapid integration of experimental data.

\langle	Ficial Bioe	energy Data	Hub														Q	¢ 🌡	Jim C
FCIC	✓ Task Portals	 Studies 	~																
ssav L	ist / EDS SEM Ima	de Analysis Ba	atches / EDS	S SEM Ima	de Analysis	Runs													
DS.	SEM Image	Δnalvs	is Resul	ts 1750				SING											
5-			is Resul		UPNENT V		CPROCES	DING											
ANAGE	ASSAY DESIGN > VIE	W BATCHES »	VIEW RUNS >	VIEW RES	ULTS VIEV	V COPY-TO-S	TUDY HISTO	RY VIEW	EXCLUDED	DATA 🕨									
=	- 🗠 - 🕹	 Copy To S 	tudy Impo	ort Data R	eplaced Filte	er 👻 🖨											1 - 15	of 15 📼	
] * [Bale 🔍 Sample ID	Image File	🗢 wt% C 🗢	wt% Fe 🔍	wt% Mg 🔍	wt% Na 🔍	wt% N 🔍	wt% 0 으	wt% Si 🔍	wt% Al 🛇	at% C 🗢	at% Fe 으	at% Me 🔍	at% Na 🔍	at% N 📀	at% O 으	at% Si 🗢	at% Al 🗢	<u>]</u>
)	2 2_1		66.85	1.53	0.19	0.05	0.92	25.03	4.45	0.98	74.93	0.37	0.1	0.03	0.88	21.06	2.13	0.49	
]	2 2_2		62.03	0.56	0.24	0.08	1.11	29.16	5.98	0.85	70.42	0.14	0.13	0.05	1.08	24.86	2.9	0.43	
)	3 3_1		49.88	2.54	0.36	0.19	1.89	28.94	13.08	3.13	61.55	0.67	0.22	0.12	2.0	26.81	6.9	1.72	
]	3 3_2		55.55	1.93	0.22	0.14	1.85	28.19	10.43	1.68	66.05	0.49	0.13	0.09	1.89	25.16	5.3	0.89	
)	3 3_3		16.01	1.6	0.0	0.0	1.43	40.01	40.69	0.26	24.58	0.53	0.0	0.0	1.89	46.11	26.72	0.18	
)	4 4_1		35.32	4.32	0.39	0.17	1.96	32.65	22.05	3.14	48.02	1.26	0.26	0.12	2.28	33.33	12.82	1.9	
]	4 4_2		49.32	3.44	0.42	0.09	3.61	33.03	8.32	1.77	59.75	0.9	0.25	0.06	3.75	30.04	4.31	0.95	
)	4 4_3		28.56	EDS_SE	M_Image	s/2_2.png	J	×	EDS	_SEM_Im	ages/4_	1.png		x 0.0	EDS_SEN	/I_Image	s/4_6.pn	g	
)	4 4_4		35.65		ALAS	9 40	100.			- 6 5	S. An	1000	101	0.0	(h)	1. area	Contraction of the	and the second second	
]	4 4_5		32.3		6.2			< / B	18		1AN	1 in	1.50	0.0	1000	- 400 -		1 00 1	
)	4 4_6		68.17	14.00				and the		-	1000	1200	d-	0.0		2	5		1 ac
כ	23 23_1	$\sim l$	36.68		a the	19.00	Charles !!	-	1	6.	and h		1 and	0.((mar)	10	m		10
1	23 23_2	1000	15.71	Alt.	200	· #	12000	-	5.00	and and a second	-	2-1-0	-	0.0	10	m s	=00		-
2	22 22 2	1000	52 98	12.		R. Songer	Sec. Sec		3	- Pain			A MA	0.(7	00		9	10
5	23 23_3	254 254	02.70	100 C			and the second se			the second second second second	A CONTRACTOR OF THE OWNER	And the second se	COLUMN AND CARD AND ADDRESS OF				Concerning of the local division of the loca	A REAL PROPERTY OF	State of the local division of the local div

Task 4 – Data Integration ²⁹

Data Finder Publications Tab



	Studies 🗸			Home Page Quality by Design Lists Fil
Welcome to the Feedstock-C	onversion Interfac	e Consortium (FCIC) 🖋	•	Data Hub Resources
		The Feedstock-Conversion Interface Consortium (FCIC) is a collaborative effort among researchers from 9 National Laboratorie Technologies Office:	s who are sponsored by the DOE's Bioenergy	Access Your Task Project Portal Find Milestone Reports
Argonne 4	SINI	Argonne National Laboratory (ANL) Los Alamos National Laboratory (LANL) Oak Ridge Nati	onal Laboratory (ORNL)	View FCIC's current Critical Property Database NEW
Laboratories CALENT OF	Idda National Schoolboy	Idaho National Laboratory (INL) National Energy Technology Laboratory (NETL) Pacific Northw	est National Laboratory (PNNL)	Please let Jim Collett or Rachel Emerson kno
V 1 2		Lawrence Berkeley National Laboratory (LBNL) National Renewable Energy Laboratory (NREL) Sandia Nationa	Laboratory (SNL)	if you have any questions or concerns.
Pacific Northwest Introduction	Los Alamos	The mission of the FCIC is develop knowledge and tools to understand and mitigate the effects of biomass feedstock and proce: consortium is organized into 8 complementary Task areas: (1) Materials of Construction; (2) Feedstock Variability; (3) Materials H Design; (5) Preprocessing; (6) High Temperature Conversion; (7) Low Temperature Conversion; and (8) Crosscutting Analyses.	s variability on biorefinery performance. The andling; (4) Data Integration for Quality by	Jim Collett Rachel Emerson james.collett@pnnl.gov 509-372-6345 208-526-1931
		The square tiles in the Data Finder tool below provide links to information, data, tools, and publications produced by Subtask Stu	dy teams within each of the FCIC Tasks. Blue	Key References
TL TECHNOLOGY		tiles indicate Subtask Studies that are associated with journal articles and reports from the FCIC that are available via the Internet the Data Finder enables you to select categories that filter the Subtask Study tiles according to your interests. The Publications	t. The faceted search tool on the left side of	FCIC FY21 Quad Charts - All Tasks
		FCIC publications.	ab takes you directly to a bibliography of	BIOENERGY TECHNOLOGIES OFFICE - 2019 R&D Sta of Technology
				BETO Organizational Chart as of 10/6/2020
Data Finder				Herbaceous Feedstock 2018 State of Technology Report
Prudias				Woody Feedstock 2018 State of Technology Report
Summary	clear all	Q Publications	Quick Help > Insert New > Manage Data >	Feedstock Supply System Design and Economics for
Publications	30			Conversion of Lignocellulosic Biomass to Hydrocarbo Fuels Conversion Pathway: Fast Pyrolysis and
Studies	20	A Review of Computational Models for the Flow of Milled Biomass Part I: Discrete-Particle Models Y Yia 1 J Stickel W Jin and J Klinner	EDIT F	Hydrotreating Bio-oil Pathway "The 2017 Design Cas
E anno		Xia, Y.; Stickel, J. J.; Jin, W.; Klinger, J., A Review of Computational Models for the Flow of Milled Biomass Part I: Discrete-Particle Models. Acs Sustain		Process Design and Economics for the Conversion of
- Status	clear	Chem Eng 2020, 8 (16), 6142-6156.		Lignocellulosic Biomass to Hydrocarbon Fuels Thermochemical Research Pathways with In Situ and
None Complete	20 / 20	VIEW DOCUMENT *		Ex Situ Upgrading of Fast Pyrolysis Vapors
Complete	30730	A Review of Computational Models for the Flow of Milled Biomass Part II: Continuum-Mechanics Models	Manuscript	LabKey Server: An open source platform for scientific
Submission Status	clear	W. Jin, J. J. Stickel, Y. Xia and J. Klinger	EDIT »	data integration, analysis and collaboration
None	0/0	Jin, W.; Stickel, J. J.; Xia, Y.; Klinger, J., A Review of Computational Models for the Flow of Milled Biomass Part II: Continuum-Mechanics Models. Acs		Understanding Pharmaceutical Quality by Design
Submitted	30 / 30	Sustain Chem Eng 2020, 6 (10), 6157-6172.		Accelerating Bioenergy Technology Advancement
	clear			Through PAIR Data Delivery
Publication Type	0/0	C A density dependent Drucker-Prager/Cap model for ring shear simulation of ground lobiolly pine	Manuscript	
Publication Type Abstract	57.2 (Fe	W. Jin, J. L. Klinger, T. L. Westover and H. Huang	EDIT +	
Publication Type Abstract Conference	0/0	Jin W: Klinger J. L: Westover T. L: Huang H. A density dependent Drucker-Prager/Cap model for ring shear simulation of ground loblolly pine. Powder		
Publication Type Abstract Conference Manuscript	0 / 0	Technolony 2020 368 45-58		
Publication Type Abstract Conference Manuscript	0 / 0 30 / 30	Technology 2020, 368, 45-58.		Attached Files
Publication Type Abstract Conference Manuscript Study	0 / 0 30 / 30 clear	Technology 2020, 368, 45-58.		Attached Files
Publication Type Abstract Conference Manuscript Study 1.1	0 / 0 30 / 30 clear 2 / 2	Technology 2020, 368, 45-58. VIEW DOCUMENT > C Advances in Multiscale Modeling of Lignocellulosic Biomass	Manuscript	Attached Files

Task 4 – Data Integration

30

Data Hub Wiki Pages



- LabKey Server provides a wide range of "webparts" that enable data owners to customize their environment.
- The screenshot to the right shows an editable Wiki webpart that allows users to integrate text and graphics into narratives to summarize research results.

CMA Identification for TCPDU Pyrolysis Performance FY19Q3 Results Summary 🖋

From the FY18 HT baseline runs, seven of the supersacks of lobiolly pine chips and residues that were preprocessed at INL and then converted to pyrolysis oils at NREL were labeled for three conversion processes efficiency metrics: 'On Stream,' a continuous variable representing the ratio of the time on stream divided by the time on stream plus the downtime; 'Char Removal', a categorical binary variable indicating whether a run needed to be stopped to clean out either cyclones or char bridging; and 'Feedtrain Bridging', a categorical binary variable indicating whether were bridging problems in the feedtrain during sample processing. The properties of the raw materials were assessed as potential explanatory variables for these three efficiency metrics including: moisture, total ash, ash speciation, elemental C, H, N, O, and S, volatiles, and particle size distribution factors (D10, D50 and D90).

There was not an obvious visual correlation or pattern between any of these factors alone and the three efficiency metrics. When using a multivariate least squares linear regression approach the 'On Stream' metric could be linearly correlated to a combination of three factors: moisture, particle size factor of D50, and total carbon. All three of these factors contribute significantly (considering p



Publications, Patents, Presentations, Awards, and Commercialization



• Collett, J., *LabKey for Multicenter R&D on Biofuels and Bio-based Products*, in *LabKey User Conference*. 2019: Seattle, WA.