



Forget Duplicating Local Changes: Apache NiFi and the Flow Development Lifecycle (FDLC)

Andy LoPresto | @yolopey

Sr. Member of Technical Staff at Hortonworks, Apache NiFi PMC & Committer

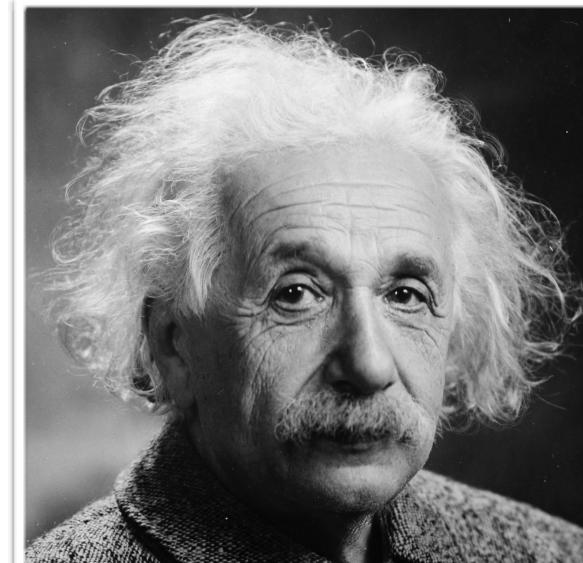
19 April 2018 Dataworks Summit Berlin

Gauging Audience Familiarity With NiFi



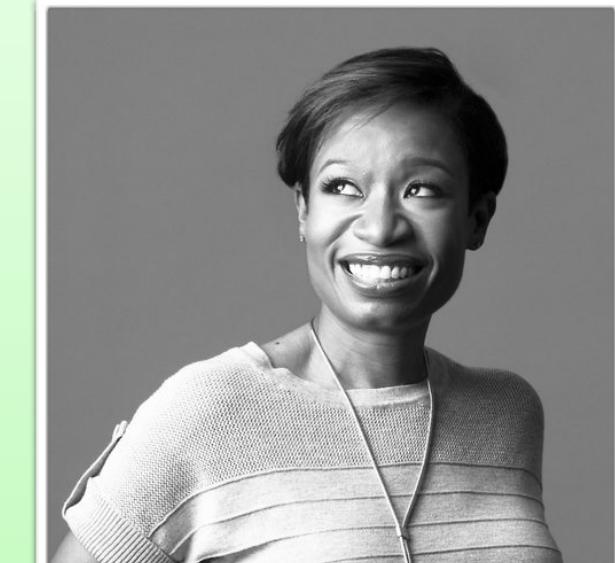
"What's a NeeFee?"

No experience with dataflow
No experience with NiFi



"I can pick this up pretty quickly"

Some experience with dataflow
Some experience with NiFi



"I refactored the Ambari integration endpoint to allow for mutual authentication TLS during my coffee break"

Forgotten more about NiFi than most of us will ever know

Agenda

- Flow development lifecycle
- Challenges in FDLC
- Apache NiFi Registry
- Roadmap



- All slides provided online, so no need to transcribe



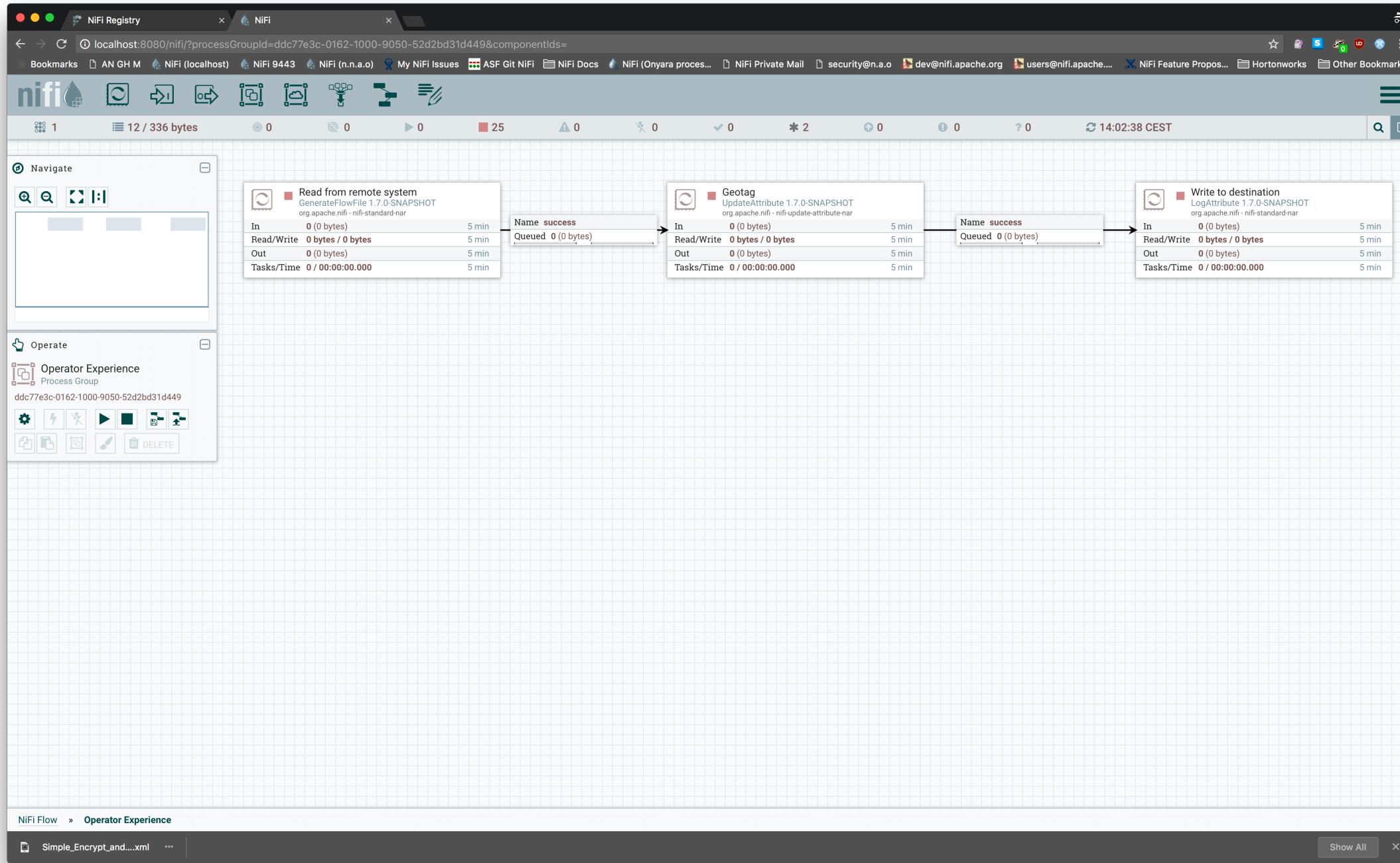
Flows Definitely Love Change

Flow Development Lifecycle (FDLC)

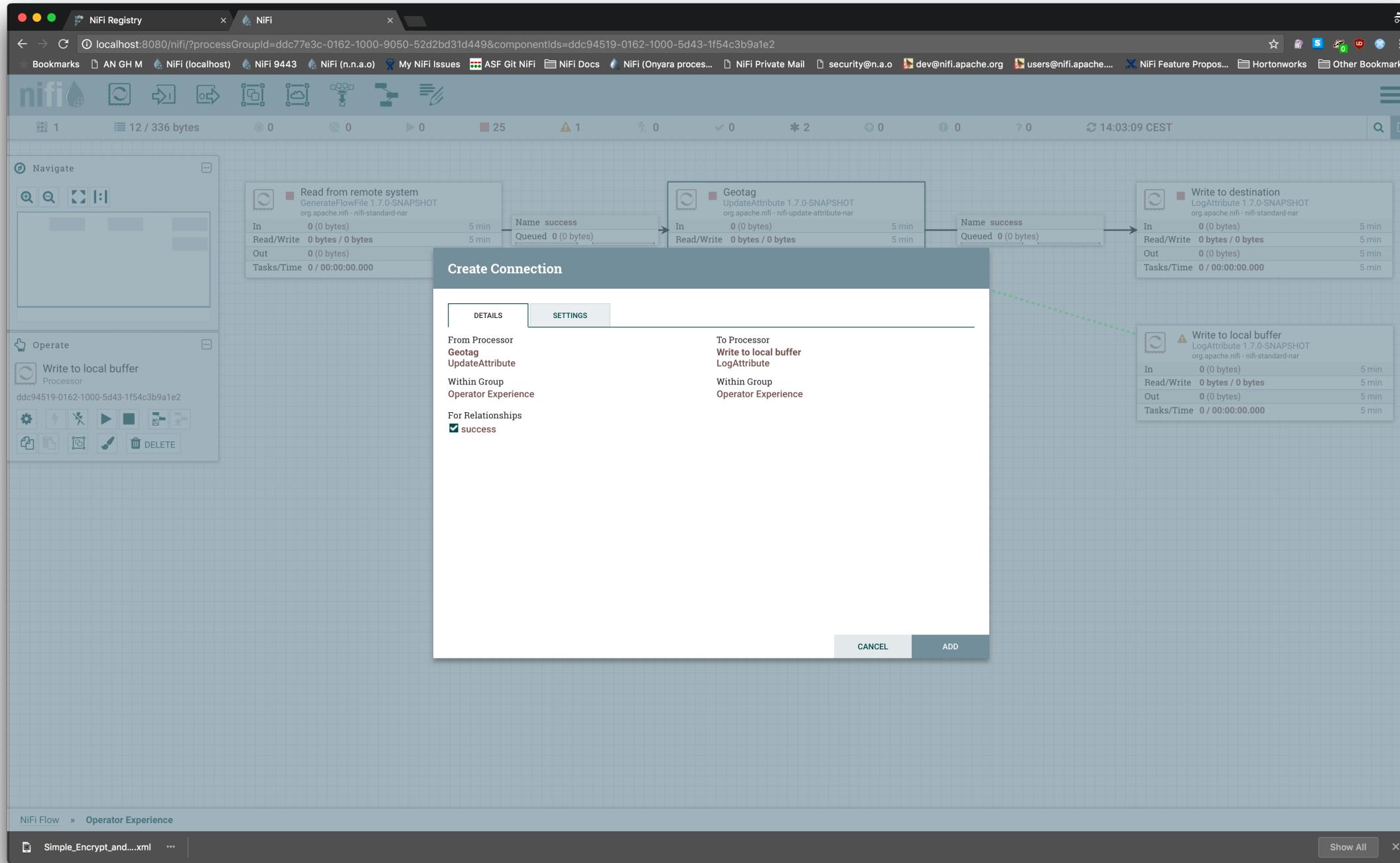
- Origins of NiFi
- Operator Experience
- MC data, don't drop, mitigate temporarily



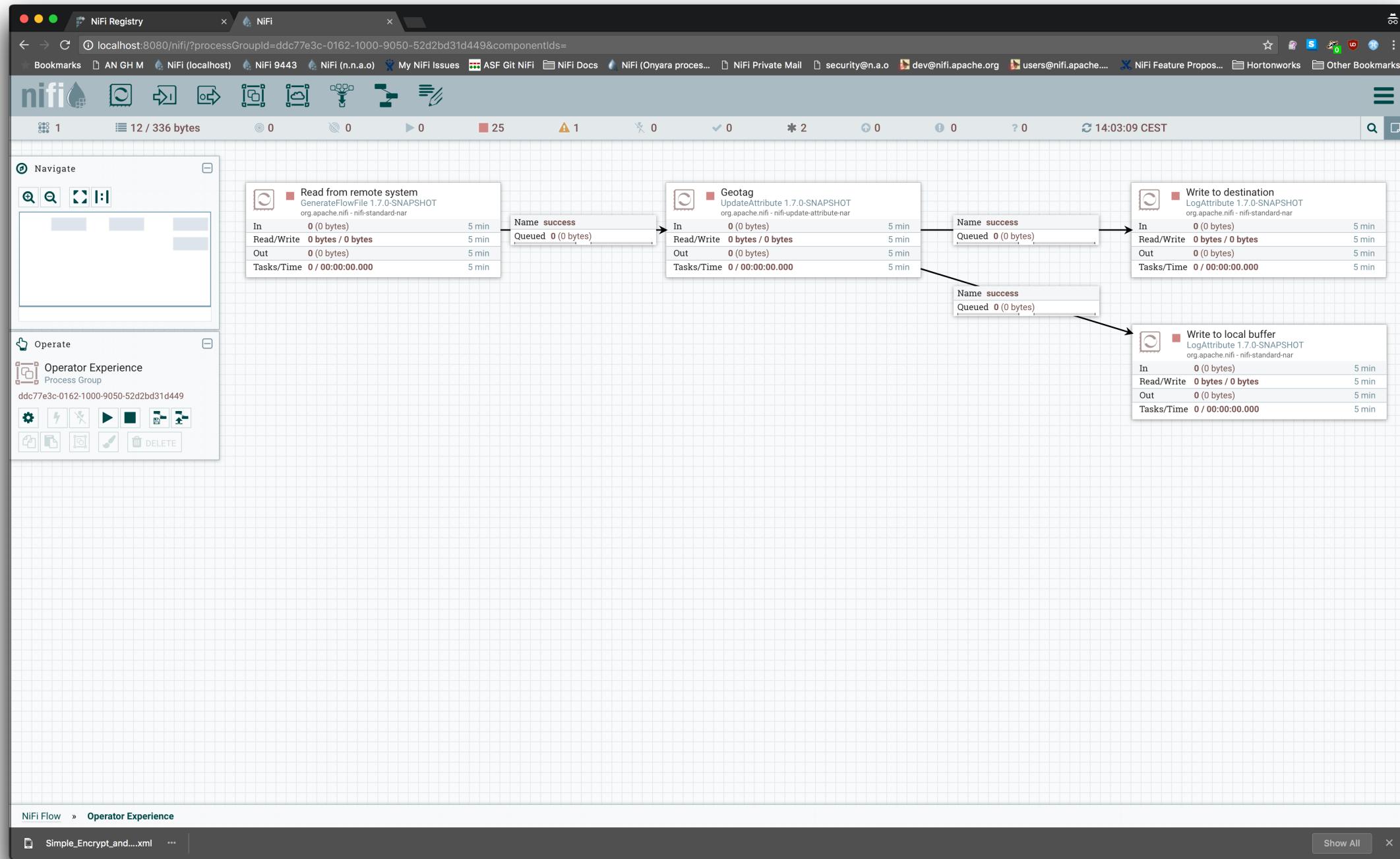
Operator Experience



Operator Experience



Operator Experience



Component Property History

- Shows previous values (user, time changed)
- Sensitive values are always encrypted at rest and never returned via the API

Configure Processor

SETTINGS SCHEDULING PROPERTIES COMMENTS

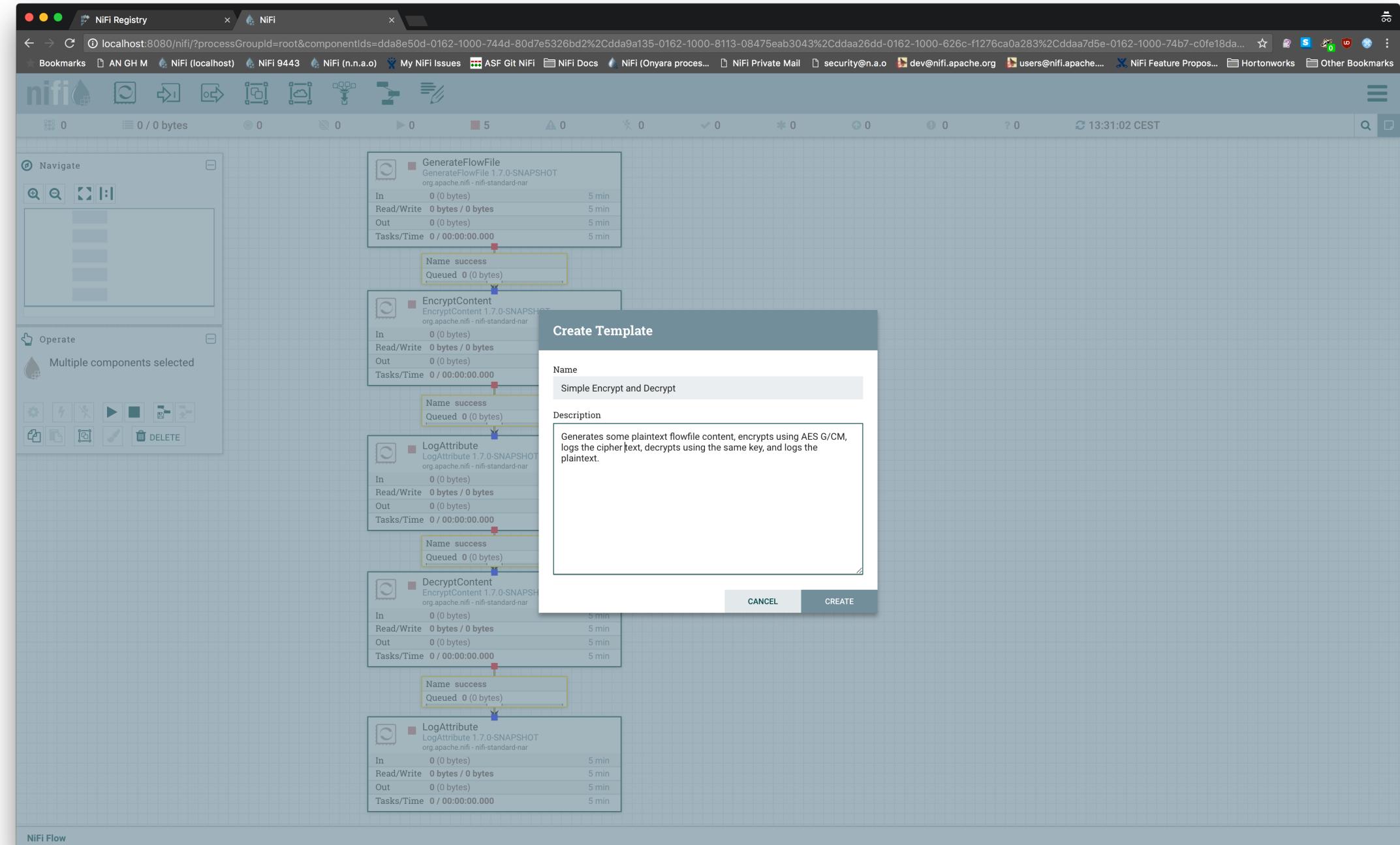
Required field +

Property	Value
Mode	Encrypt
Key Derivation Function	None
Encryption Algorithm	AES_GCM
Allow insecure cryptographic modes	Not Allowed
Password	No value set
Raw Key (hexadecimal)	Sensitive value set
Public Keyring File	In keyed encryption, this is the raw key, encoded in hexadecimal
Public Key User Id	Supports expression language: false
Private Keyring File	History: • ***** - 02/21/2017 20:30:46 PST (CN=alopresto, OU=Apache NiFi)
Private Keyring Passphrase	No value set No value set

CANCEL APPLY

Exporting Flows

- XML templates
- Copying flow.xml.gz between systems



Flow Developers Lose Consciousness

Challenges

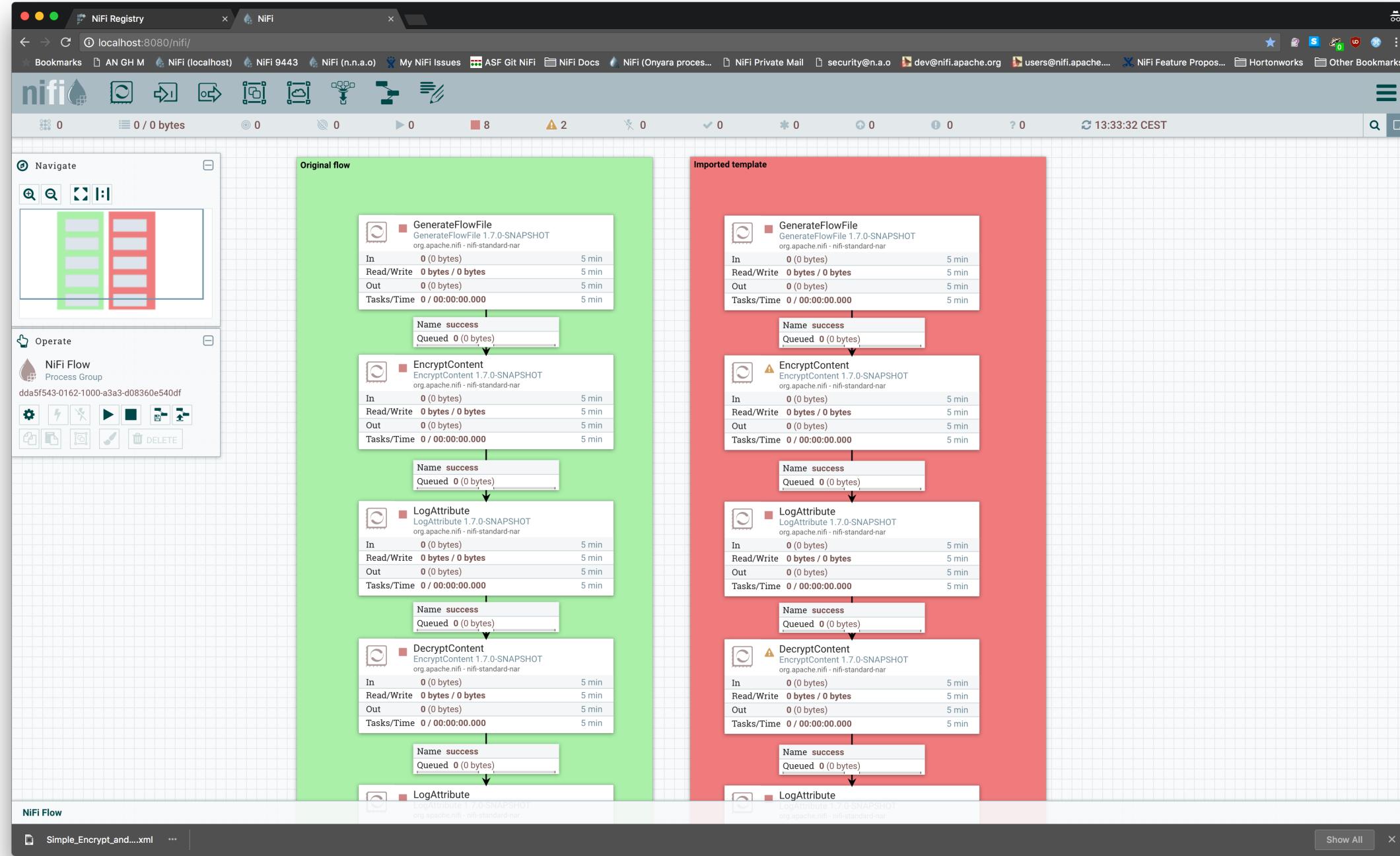
- Templates
 - Updates/replacement
 - Sensitive property replacement
- Flow.xml.gz migration
 - Key synchronization
- Environment promotion
 - Approval processes
 - Verifiability

Template Replacement

- Export a new version of template
- Transfer (somehow)
- Verify?
- Import onto canvas side-by-side existing flow
- Stop processors
- Empty queues
- Reconnect queues
- Start
- Pray?

```
-rw-r--r-- 1 alopresto staff 381B Apr 19 15:02 flow_20170626-171416_label_description.xml  
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170628-111620_xxe_transform_xml.xml  
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170628-155452_xxe_with_patch.xml  
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170629-103436_xxe_password_leak.xml  
-rw-r--r-- 1 alopresto staff 1.9K Apr 19 15:02 flow_20170717-192339_trusted_hostname.xml  
-rw-r--r-- 1 alopresto staff 2.7K Apr 19 15:02 flow_20170718-101751_evaluate_json_path.xml  
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170721-175745_site_to_site_secure.xml  
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170724-161539.xml  
-rw-r--r-- 1 alopresto staff 5.7K Apr 19 15:02 flow_20170724-181056_listen_http_sslv3.xml  
-rw-r--r-- 1 alopresto staff 2.3K Apr 19 15:02 flow_20170726-155905_encrypted_spk.xml  
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170802-172619_simple_gen_and_log.xml  
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170804-164011_simple_gen_and_log.xml  
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20170807-154118_groovy_json_uppercase.xml  
-rw-r--r-- 1 alopresto staff 1.8K Apr 19 15:02 flow_20170807-220536_put_mongo_record.xml  
-rw-r--r-- 1 alopresto staff 6.1K Apr 19 15:02 flow_20170809-105304_date_difference.xml  
-rw-r--r-- 1 alopresto staff 6.4K Apr 19 15:02 flow_20170809-174400_json_record_path.xml  
-rw-r--r-- 1 alopresto staff 1.3K Apr 19 15:02 flow_20170809-210516_validate_xml.xml  
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170815-102522_json_to_sql.xml  
-rw-r--r-- 1 alopresto staff 1.5K Apr 19 15:02 flow_20170822-104407_json_formatter.xml  
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20170823-123729_evaluatexpath.xml  
-rw-r--r-- 1 alopresto staff 2.5K Apr 19 15:02 flow_20170825-115758_encrypt_and_decrypt.xml  
-rw-r--r-- 1 alopresto staff 2.0K Apr 19 15:02 flow_20170825-121351_encrypt_decrypt_clean.xml  
-rw-r--r-- 1 alopresto staff 2.0K Apr 19 15:02 flow_20170825-121656_encrypt_decrypt_aligned.xml  
-rw-r--r-- 1 alopresto staff 2.3K Apr 19 15:02 flow_20170825-190225_listen_http_and_invoke_http_with_tls1_2.xml  
-rw-r--r-- 1 alopresto staff 5.0K Apr 19 15:02 flow_20170831-153220_psaltis_meetup_template_local.xml  
-rw-r--r-- 1 alopresto staff 4.0K Apr 19 15:02 flow_20170905-175343_xxe_template.xml  
-rw-r--r-- 1 alopresto staff 4.3K Apr 19 15:02 flow_20170905-185907_funnel_bus.xml  
-rw-r--r-- 1 alopresto staff 8.1K Apr 19 15:02 flow_20170906-111801_chained_get_html_element_NIFI-4356.xml  
-rw-r--r-- 1 alopresto staff 2.5K Apr 19 15:02 flow_20170909-121848_sydney_demo.xml  
-rw-r--r-- 1 alopresto staff 3.2K Apr 19 15:02 flow_20170919-202215_meetup_sydney.xml  
-rw-r--r-- 1 alopresto staff 3.8K Apr 19 15:02 flow_20170920-190328_crash_course.xml  
-rw-r--r-- 1 alopresto staff 2.9K Apr 19 15:02 flow_20170926-180548_handle_http_with_tls1_2.xml  
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20170929-172333_encrypt_decrypt_1_4_0_verification.xml  
-rw-r--r-- 1 alopresto staff 6.3K Apr 19 15:02 flow_20171003-192928_update_record_and_replace_text.xml  
-rw-r--r-- 1 alopresto staff 8.9K Apr 19 15:02 flow_20171004-180529_web_server_and_json.xml  
-rw-r--r-- 1 alopresto staff 3.8K Apr 19 15:02 flow_20171103-183334_groovy_json_to_sql_srsw.xml  
-rw-r--r-- 1 alopresto staff 1.5K Apr 19 15:02 flow_20171127-190651_local_site_to_site.xml  
-rw-r--r-- 1 alopresto staff 2.1K Apr 19 15:02 flow_20180101-145059_registry_rc_verification.xml  
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20180104-104211_test_count_text.xml  
-rw-r--r-- 1 alopresto staff 2.0K Apr 19 15:02 flow_20180104-112138_test_count_text_remote.xml  
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-100218_consume_jms_test.xml  
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-105514.xml  
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-131629_consume_jms_10_threads.xml  
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-160620_consume_jms_multithread.xml  
-rw-r--r-- 1 alopresto staff 1.6K Apr 19 15:02 flow_20180208-200731_simple_demo_route_on_time_parity.xml  
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20180214-182110_jms_deserialization_test.xml  
-rw-r--r-- 1 alopresto staff 2.6K Apr 19 15:02 flow_20180216-205336_attribute_evaluator.xml  
-rw-r--r-- 1 alopresto staff 5.7K Apr 19 15:02 flow_20180228-130457_attribute_sqo.xml  
-rw-r--r-- 1 alopresto staff 8.3K Apr 19 15:02 flow_20180307-171740_NIFI-4928.xml  
-rw-r--r-- 1 alopresto staff 1.5K Apr 19 15:02 flow_20180313-171310_python_process_attribute.xml  
-rw-r--r-- 1 alopresto staff 2.4K Apr 19 15:02 flow_20180315-113634_NIFI-4246-OAuth.xml  
-rw-r--r-- 1 alopresto staff 1.9K Apr 19 15:02 flow_20180328-115056.xml  
-rw-r--r-- 1 alopresto staff 2.5K Apr 19 15:02 flow_20180405-111419_prioritizers.xml
```

Template Replacement





Finally, Dataflow Lessons Conquer

NiFi Registry for Dataflows

Introducing Apache NiFi Registry 0.1.0

- Previously, flows were exported via XML templates
 - Didn't contain sensitive values
 - Couldn't be updated in-place
 - No tracking system
- NiFi Registry brings asset management as first-class citizen to NiFi
- Flows can be versioned
- Flows can be promoted between environments

The screenshot shows the Apache NiFi Registry interface. At the top, there's a navigation bar with a logo, the text "NiFi Registry / All", and user information like "registry_user" and "LOGOUT". Below the header, there's a search bar and sorting options ("Sort by: Name (a-z)"). The main area displays two flows: "Flow 1 - Bucket 1" (1 version) and "Flow 2 - Bucket 2" (2 versions). For "Flow 2", the second version is expanded, showing a "DESCRIPTION" field with "Description 2" and a "CHANGE LOG" section. The change log lists:

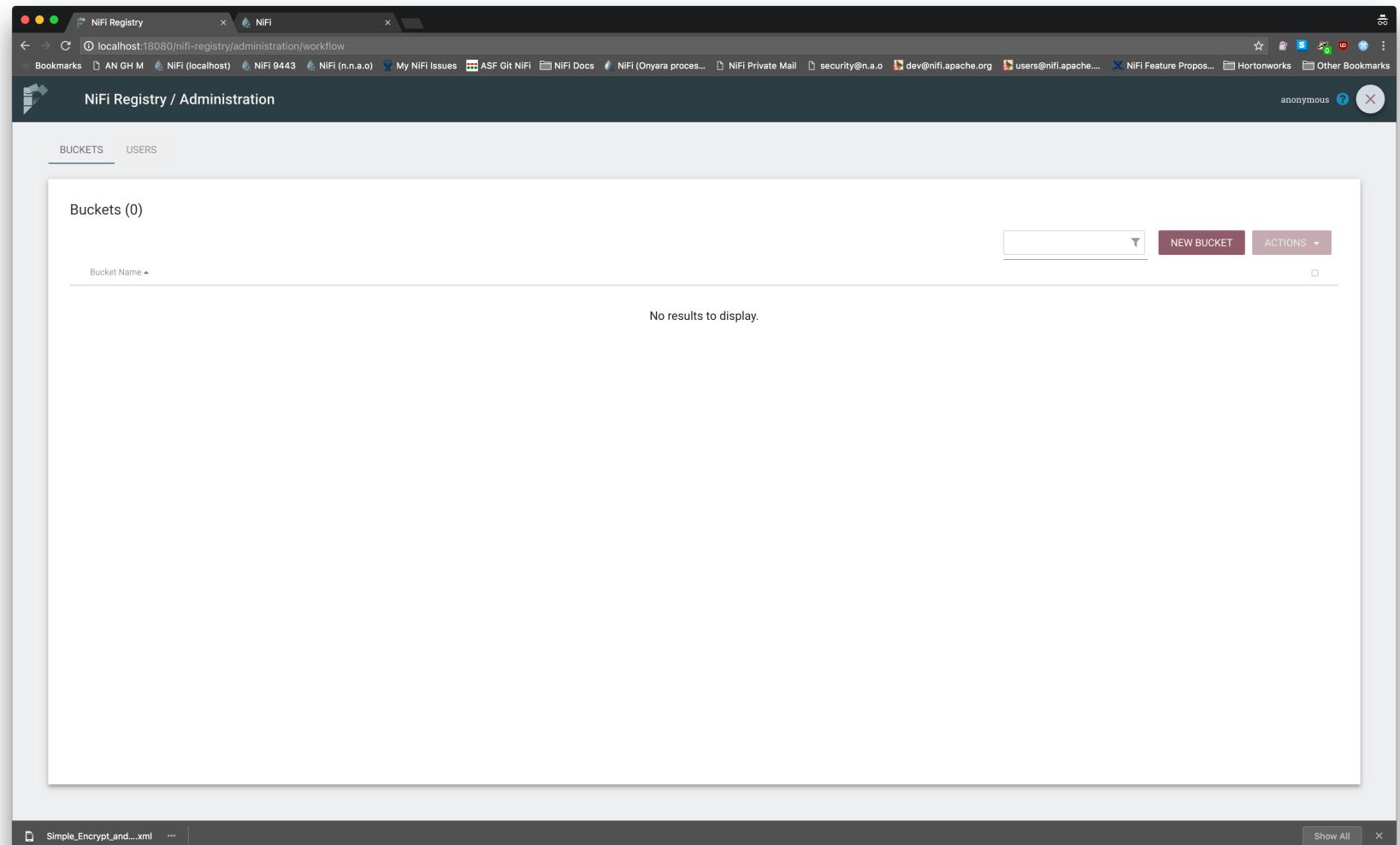
- Version 2 - 40 minutes ago by registry_user: Add processors (Dec-26-2017 at 11:23 PM)
- Version 1 - 41 minutes ago by registry_user



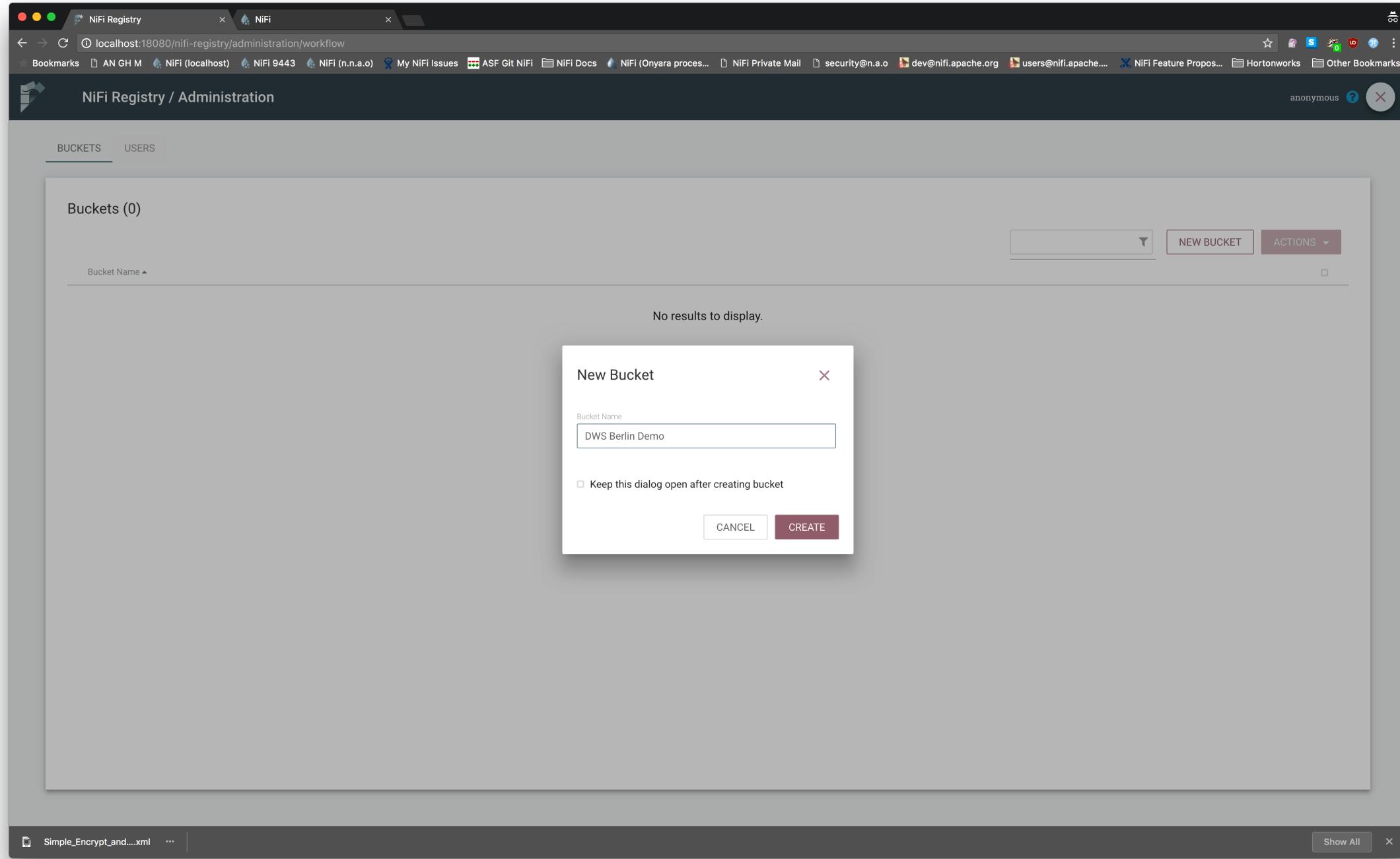
First, Demos Look Cool

Create Registry

- Install nifi-registry
- \$ mvn clean install
- \$./bin/nifi-registry.sh start
- Browse to <http://localhost:18080>



Create Bucket



Connect to NiFi

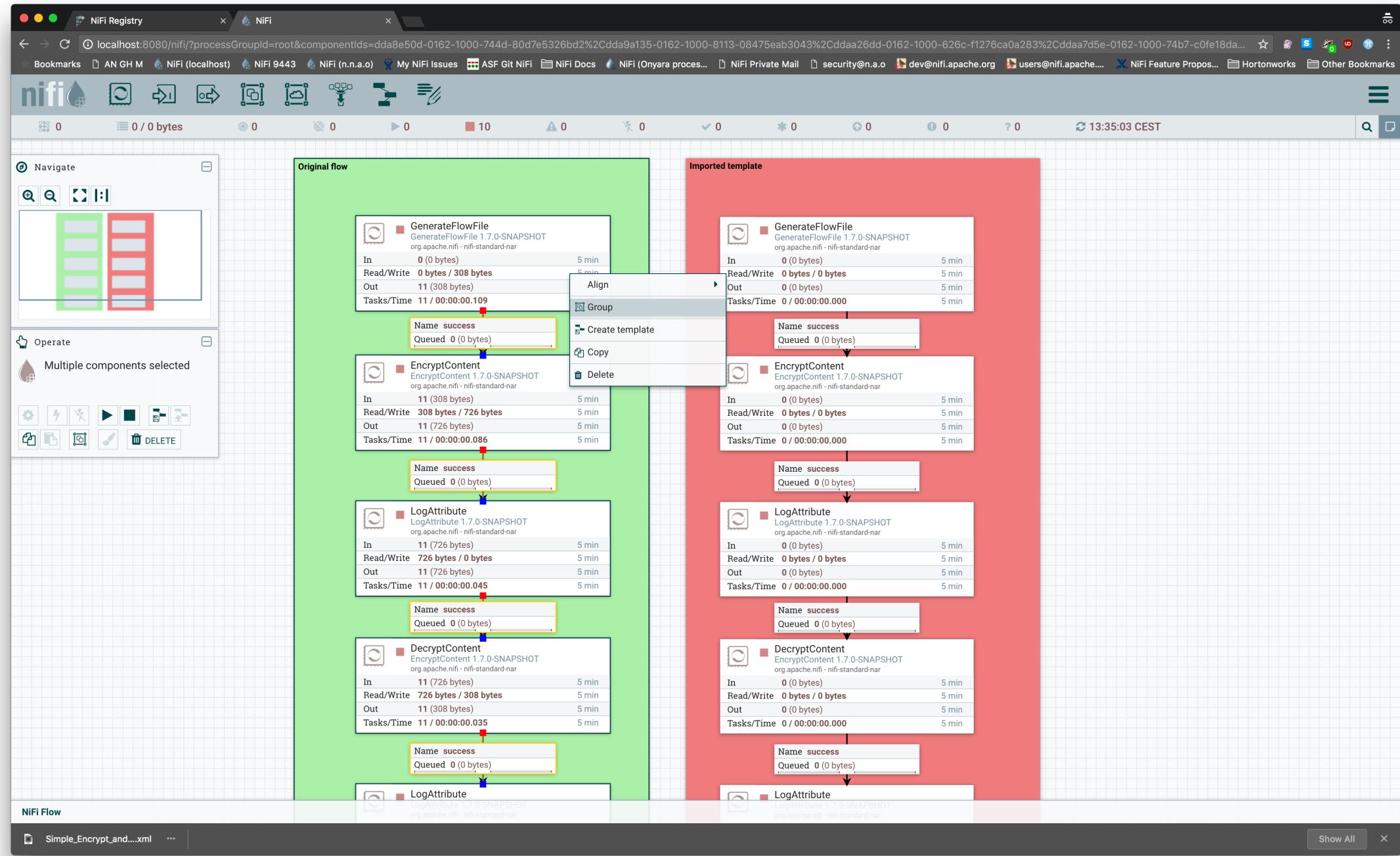
The screenshot shows a web browser window for the NiFi Registry at localhost:8080/nifi/. The main page displays 'NiFi Settings' with tabs for GENERAL, REPORTING TASK CONTROLLER SERVICES, REPORTING TASKS, and REGISTRY CLIENTS. The REGISTRY CLIENTS tab is active, showing a table with columns: Name, Location, and Description. A modal dialog titled 'Add Registry Client' is open in the foreground, prompting for 'Name' (DWS Berlin Registry), 'URL' (http://localhost:18080), and 'Description' (A NiFi Registry instance used during the DWS Berlin FDLC demo.). At the bottom of the dialog are 'CANCEL' and 'ADD' buttons.

Connect to NiFi

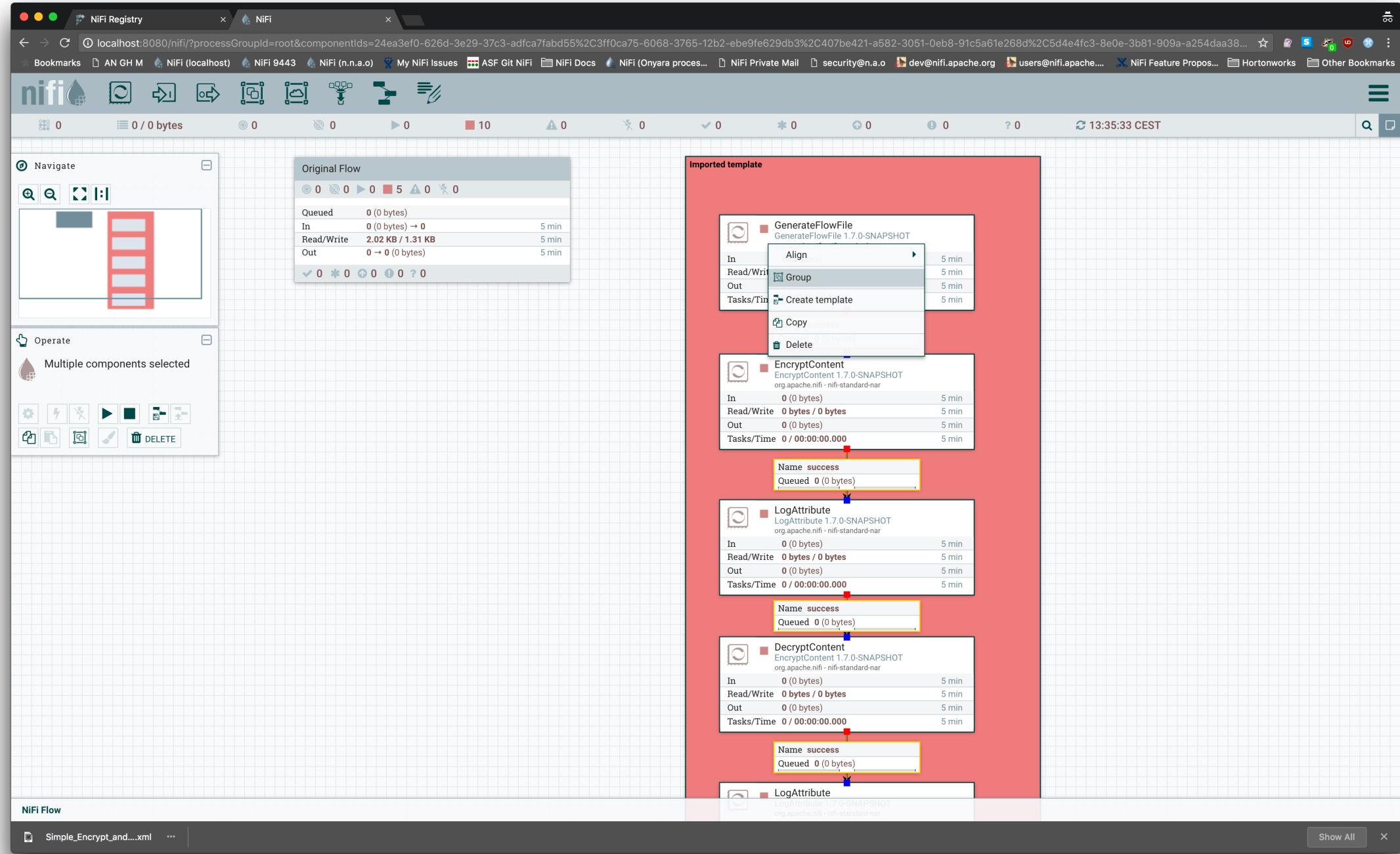
The screenshot shows the NiFi Registry interface in a web browser. The title bar indicates the window is titled "NiFi Registry" and "NiFi". The address bar shows the URL "localhost:8080/nifi/". The main content area is titled "NiFi Settings" and displays a table of "REGISTRY CLIENTS". The table has columns for "Name", "Location", and "Description". There is one entry: "DWS Berlin Registry" with "http://localhost:18080" as the location and a descriptive note about it being used during a demo. The bottom of the screen shows a "NIFI FLOW" tab with a file named "Simple_Encrypt_and....xml".

Name	Location	Description
DWS Berlin Registry	http://localhost:18080	A NiFi Registry instance used during the DWS Berlin FDLC demo.

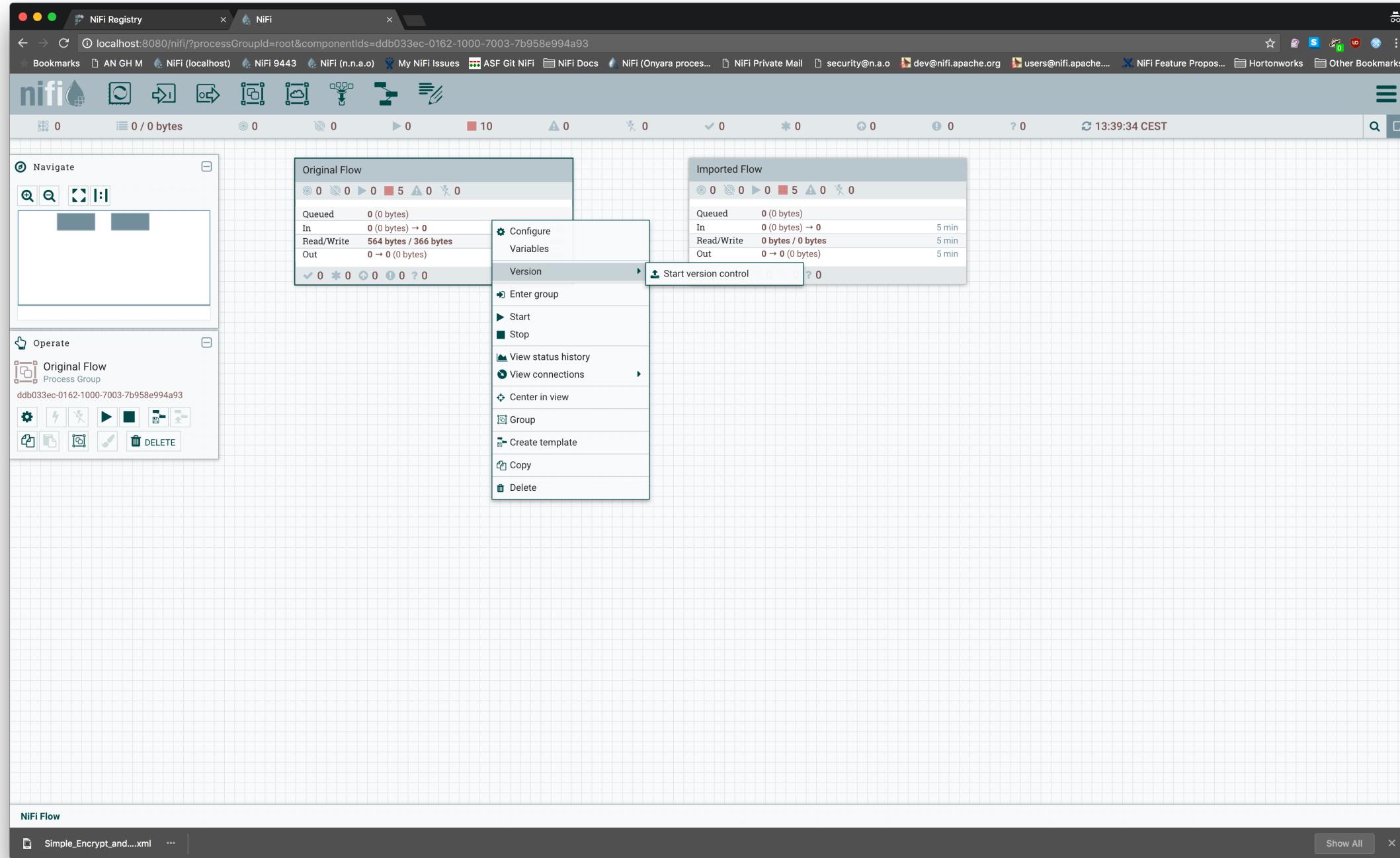
Create Process Group



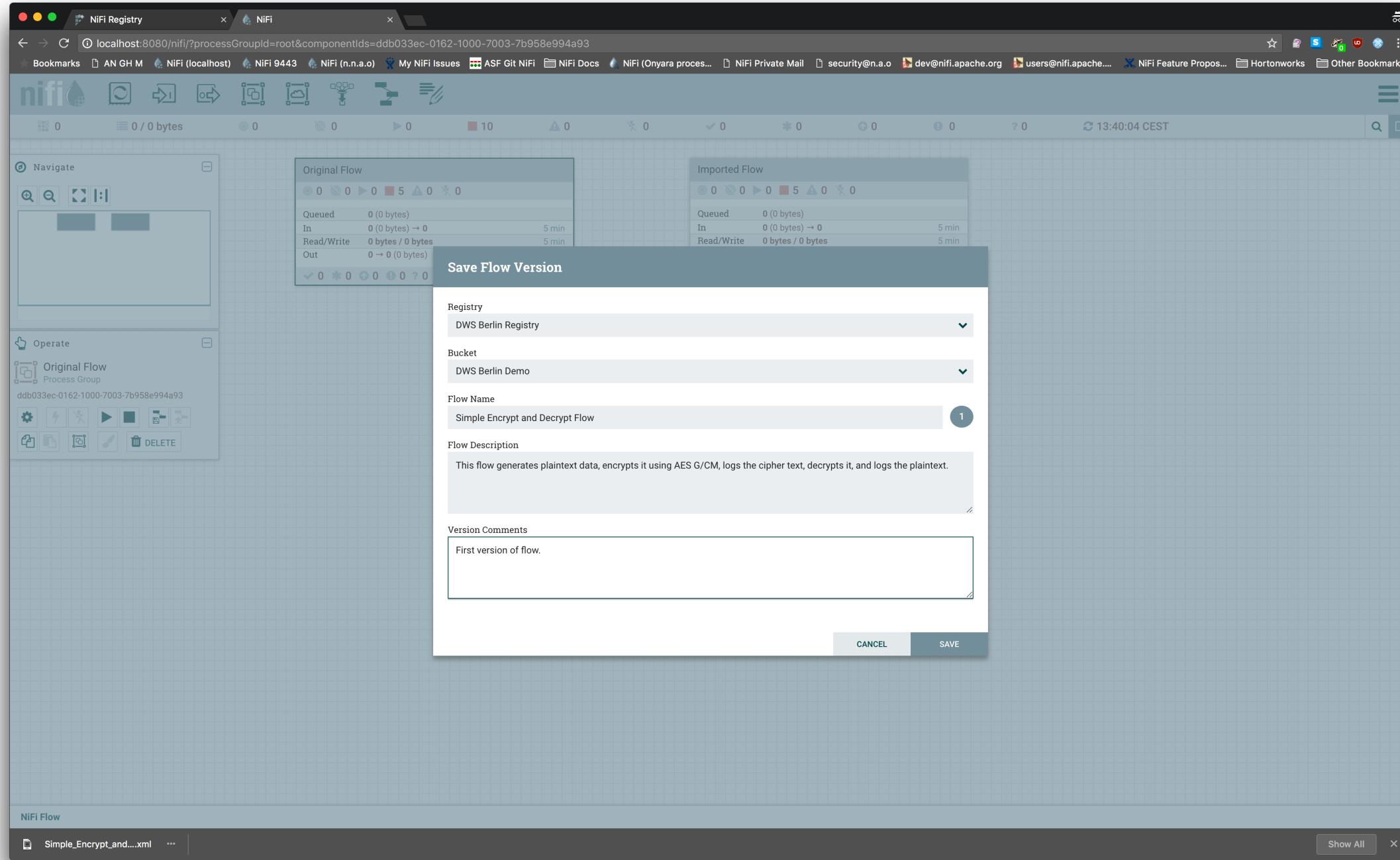
Create Process Group



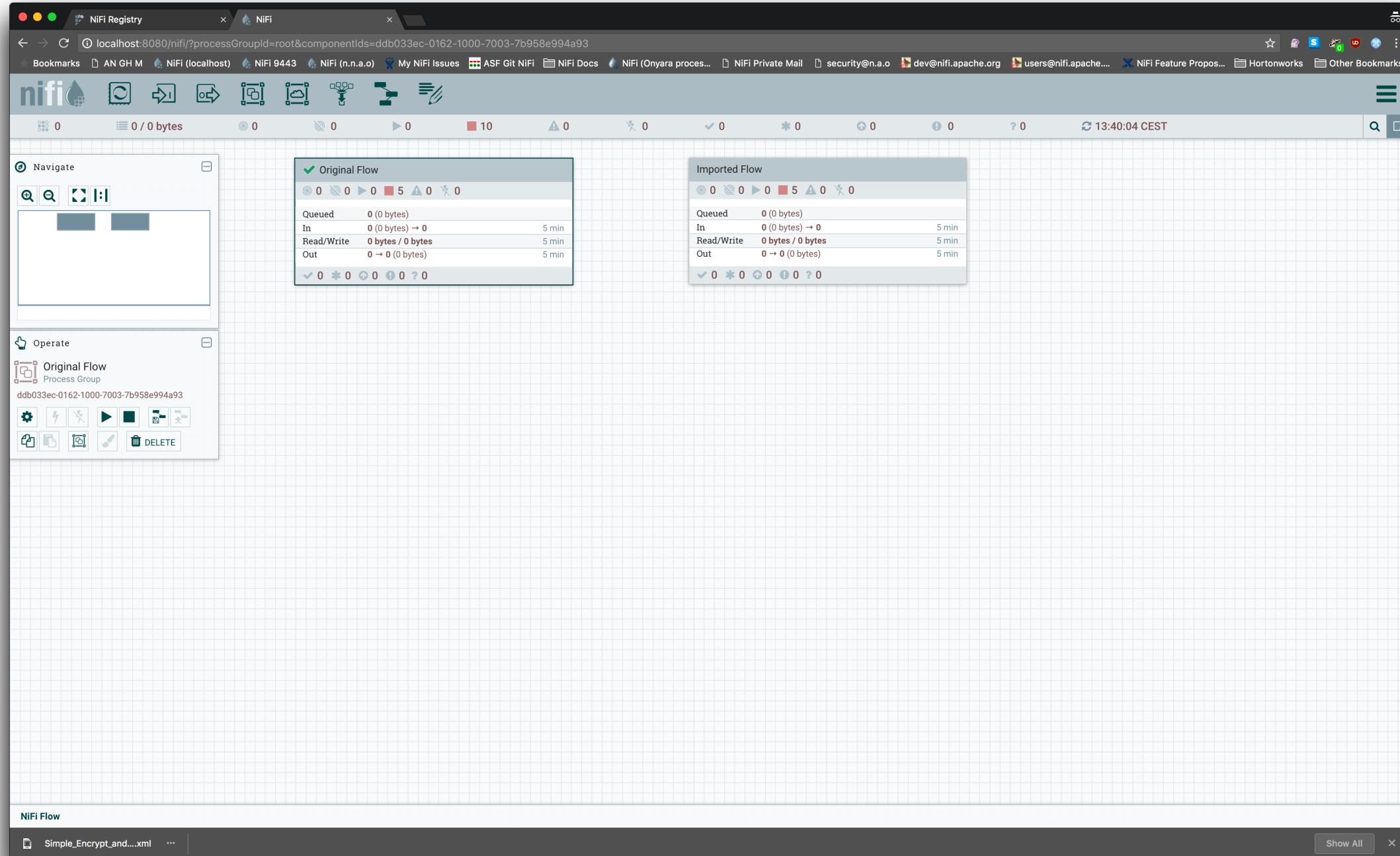
Commit Version



Commit Version



Commit Version



View flow in Registry

The screenshot shows a web browser window displaying the NiFi Registry at localhost:18080/nifi-registry/explorer/grid-list. The title bar of the browser says "NiFi Registry" and "NiFi". The address bar shows the URL. The page header includes a "NiFi Registry / All" dropdown, a user status "anonymous", and a "Sort by: Name (a - z)" dropdown.

The main content area displays a flow named "Simple Encrypt And Decrypt Flow - DWS Berlin Demo". It shows "2 VERSIONS".

BUCKET IDENTIFIER: d7df1fc1-d14b-4bb5-a8a6-ea57a60c5c98

FLOW IDENTIFIER: b288f8eb-8adf-455a-b887-785239e97d33

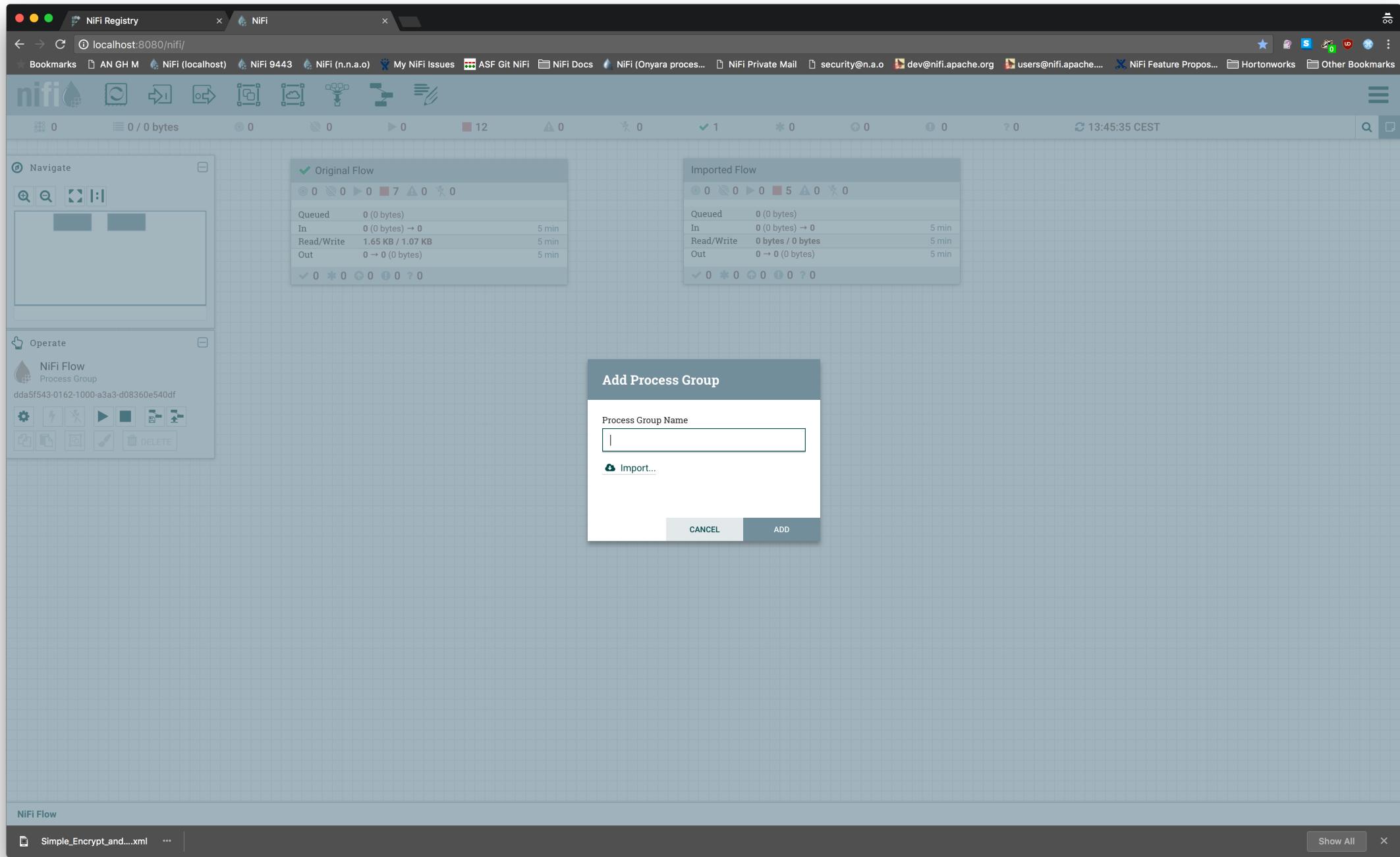
DESCRIPTION: This flow generates plaintext data, encrypts it using AES GCM, logs the cipher text, decrypts it, and logs the plaintext.

CHANGE LOG:

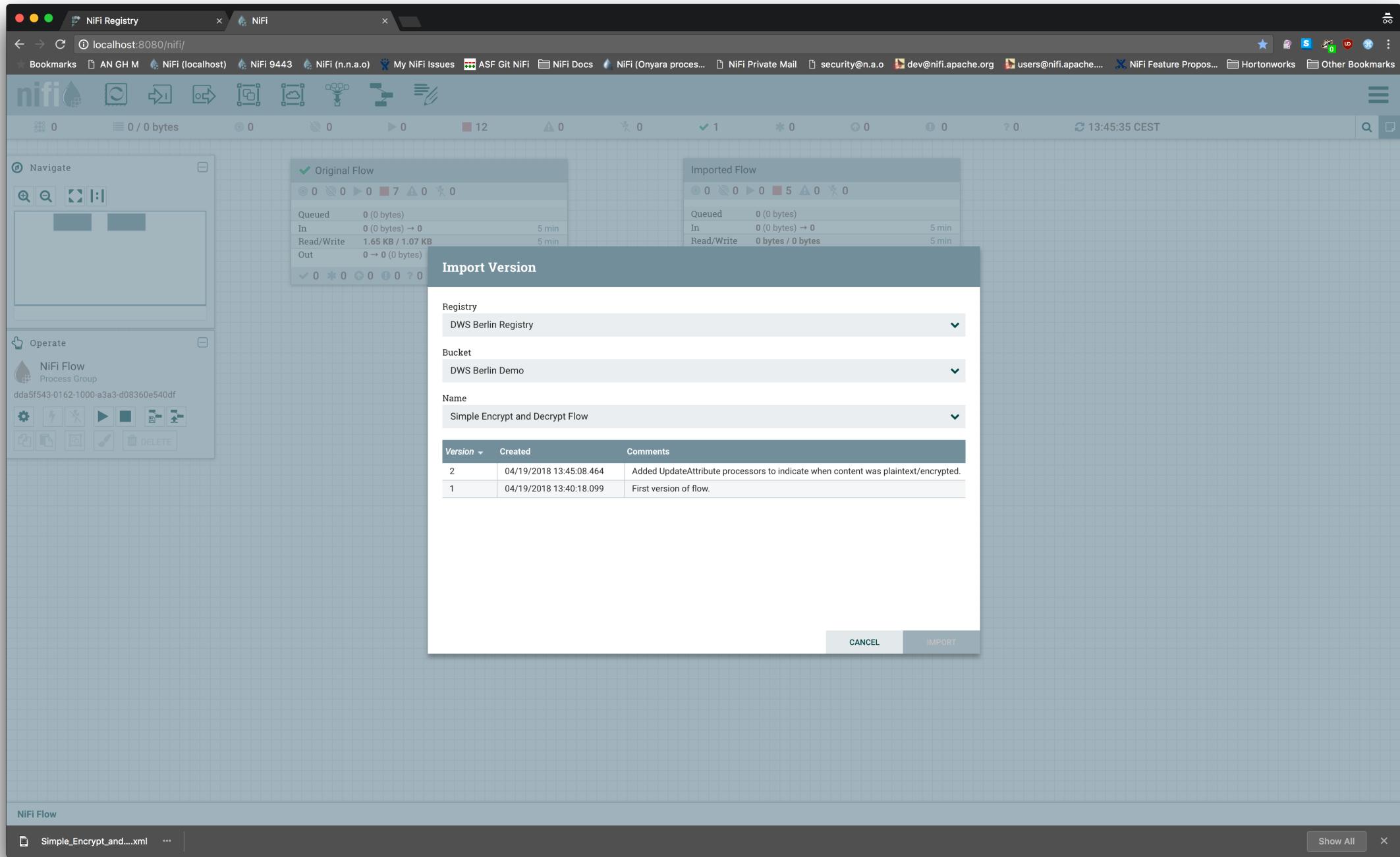
- Version 2 - a few seconds ago by anonymous:
Added UpdateAttribute processors to indicate when content was plaintext/encrypted.
Apr-19-2018 at 1:45 PM
- Version 1 - 5 minutes ago by anonymous

At the bottom of the browser window, there is a tab labeled "Simple_Encrypt_and....xml" and a "Show All" button.

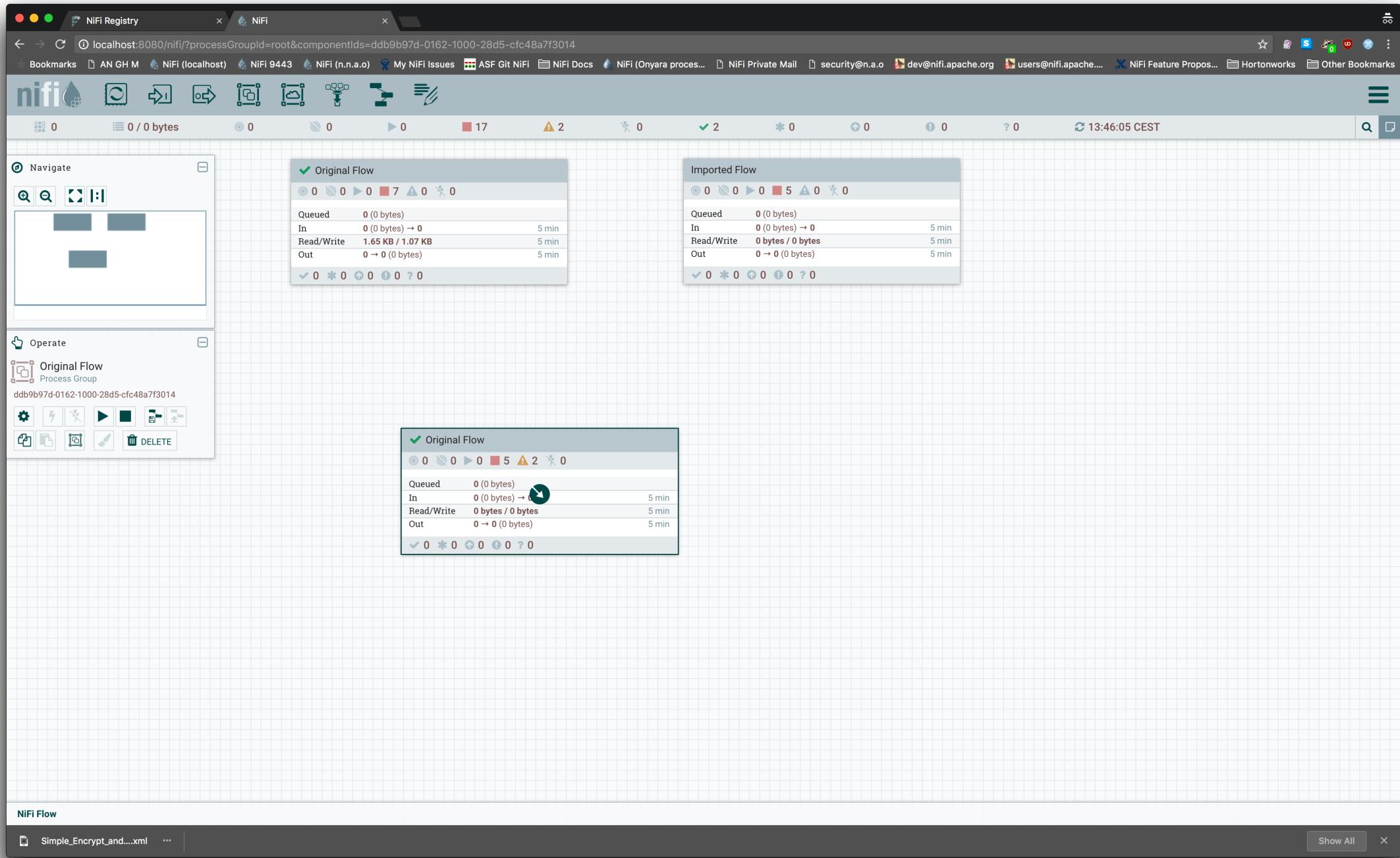
Import new instance into NiFi



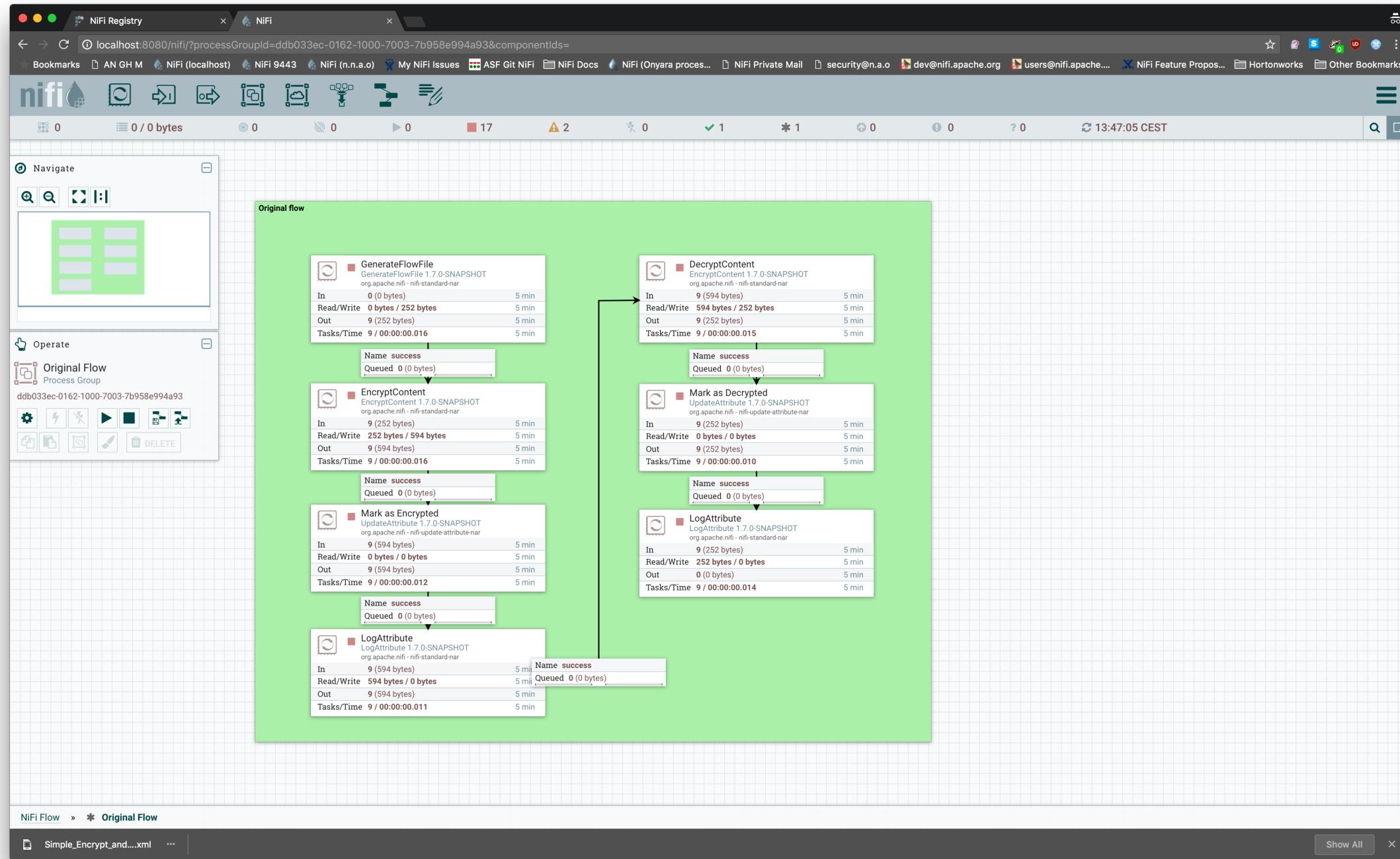
Import new instance into NiFi



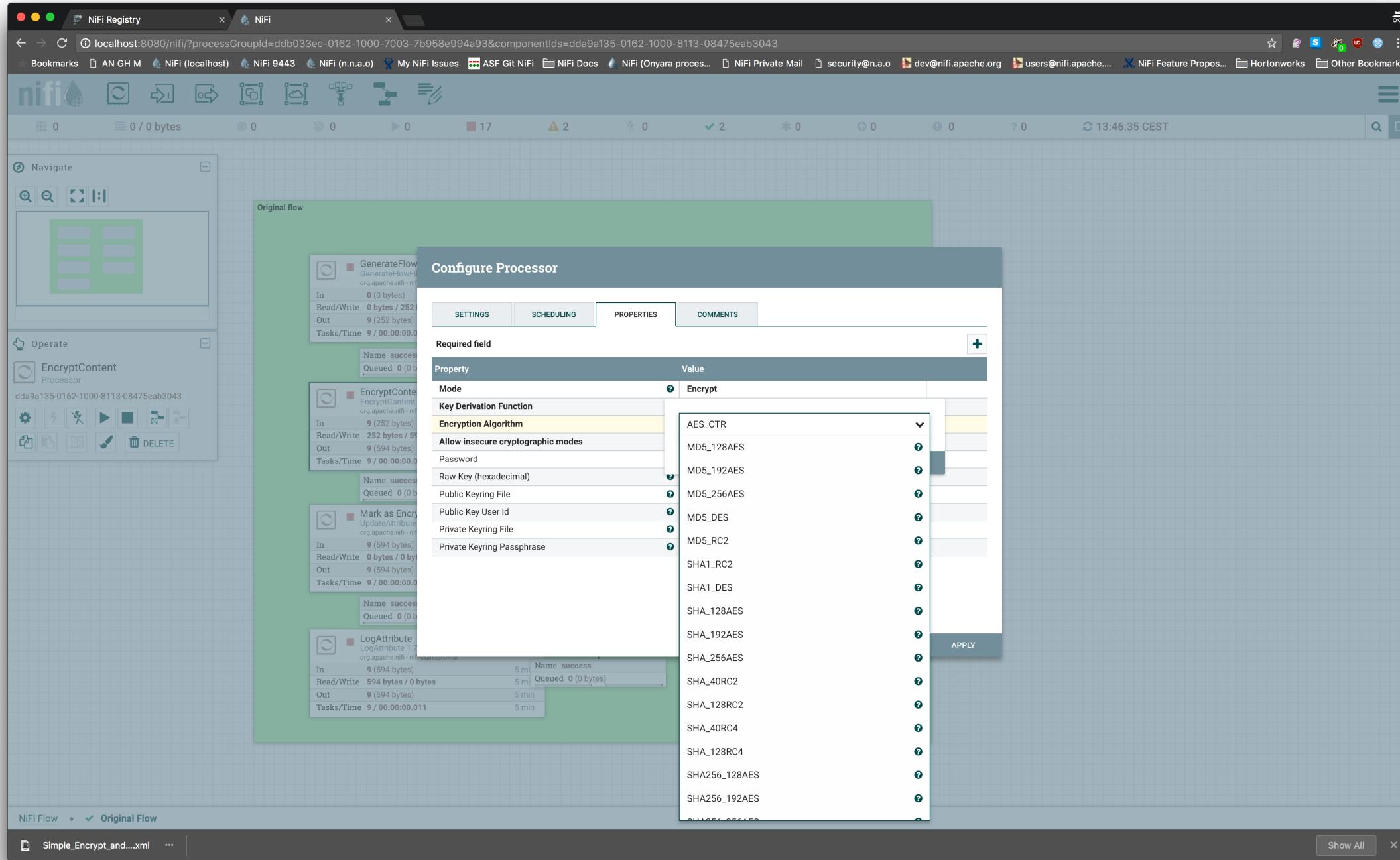
Import new instance into NiFi



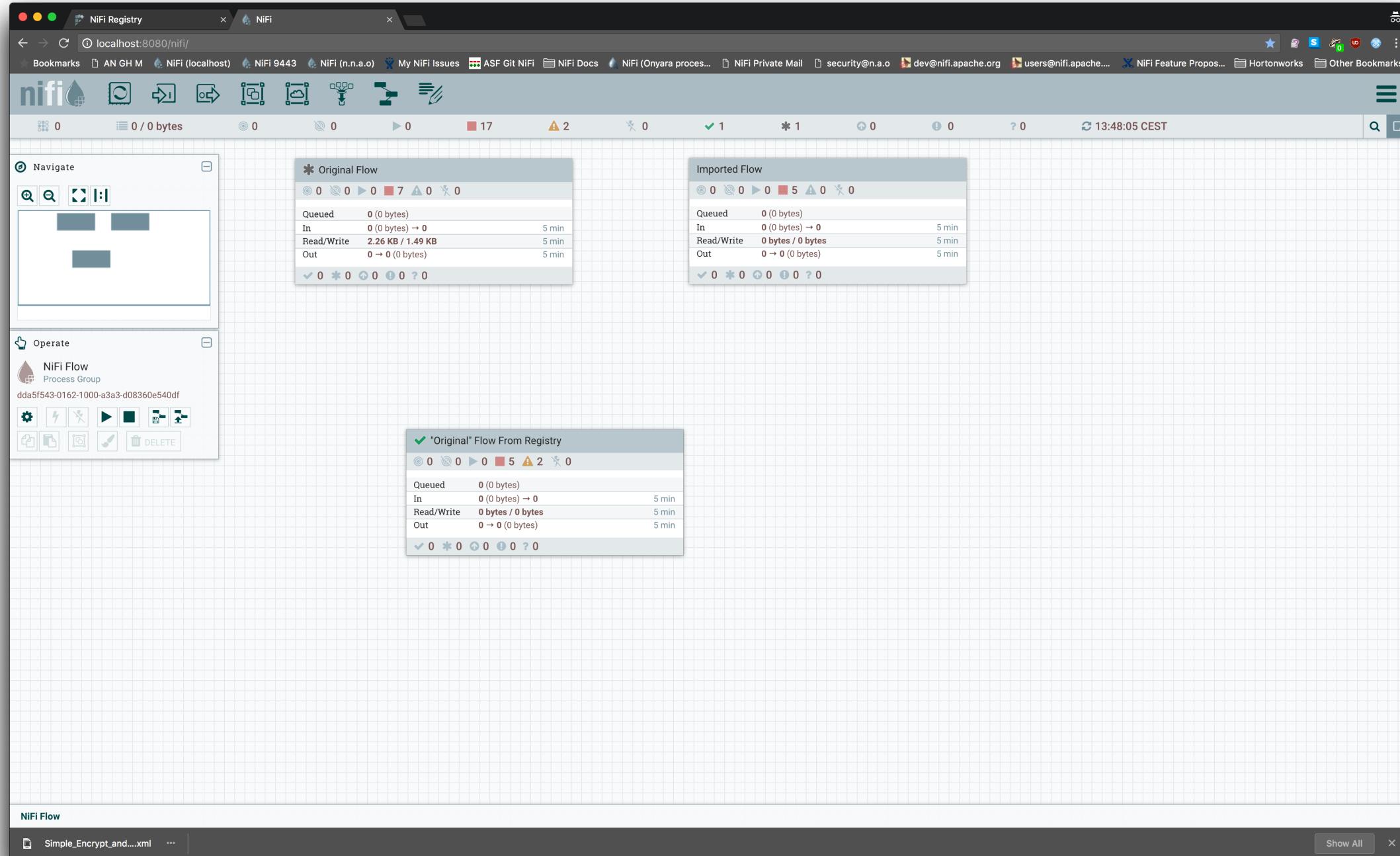
Modify the original flow



Modify the original flow

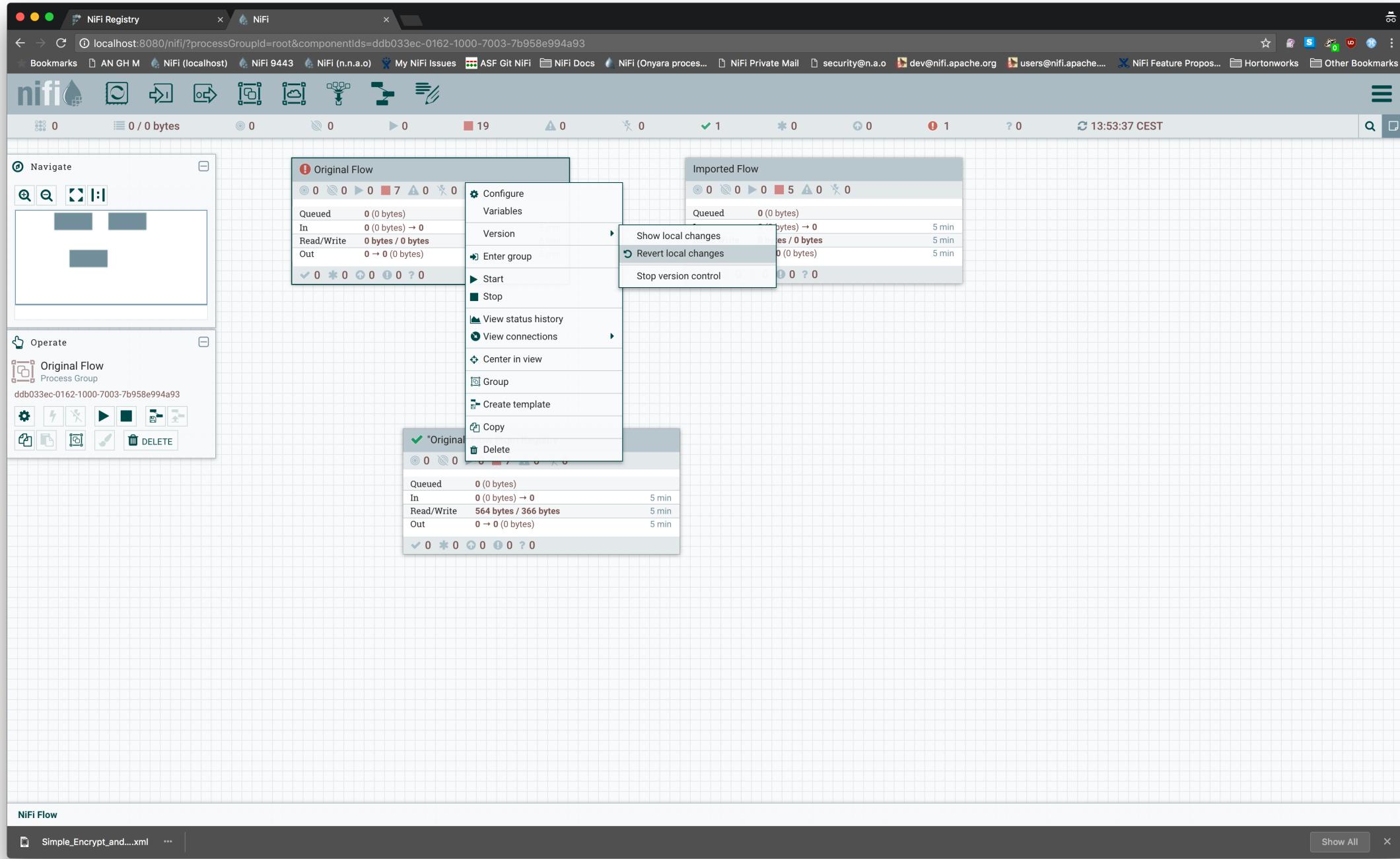


Modify the original flow

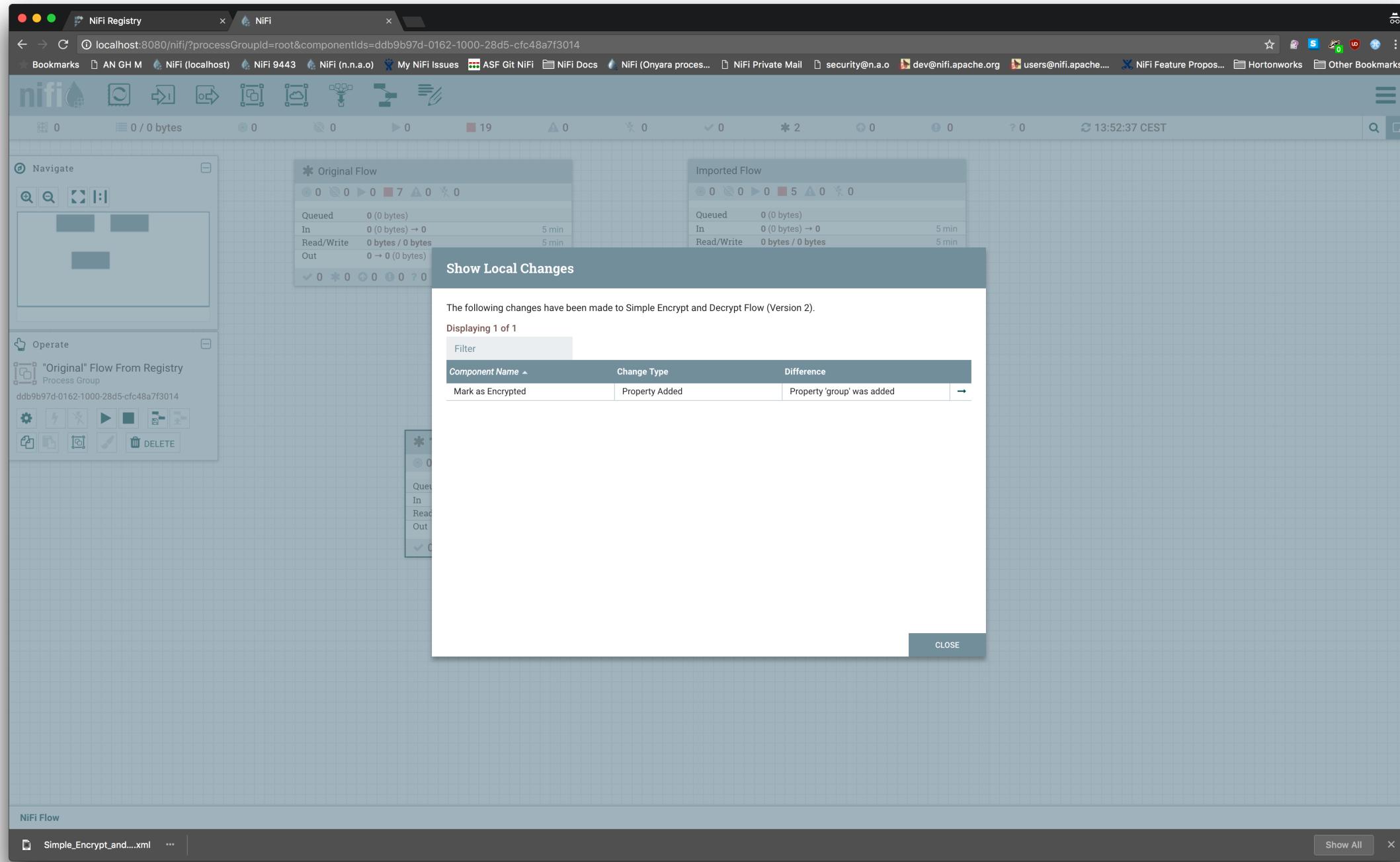


See local changes before committing

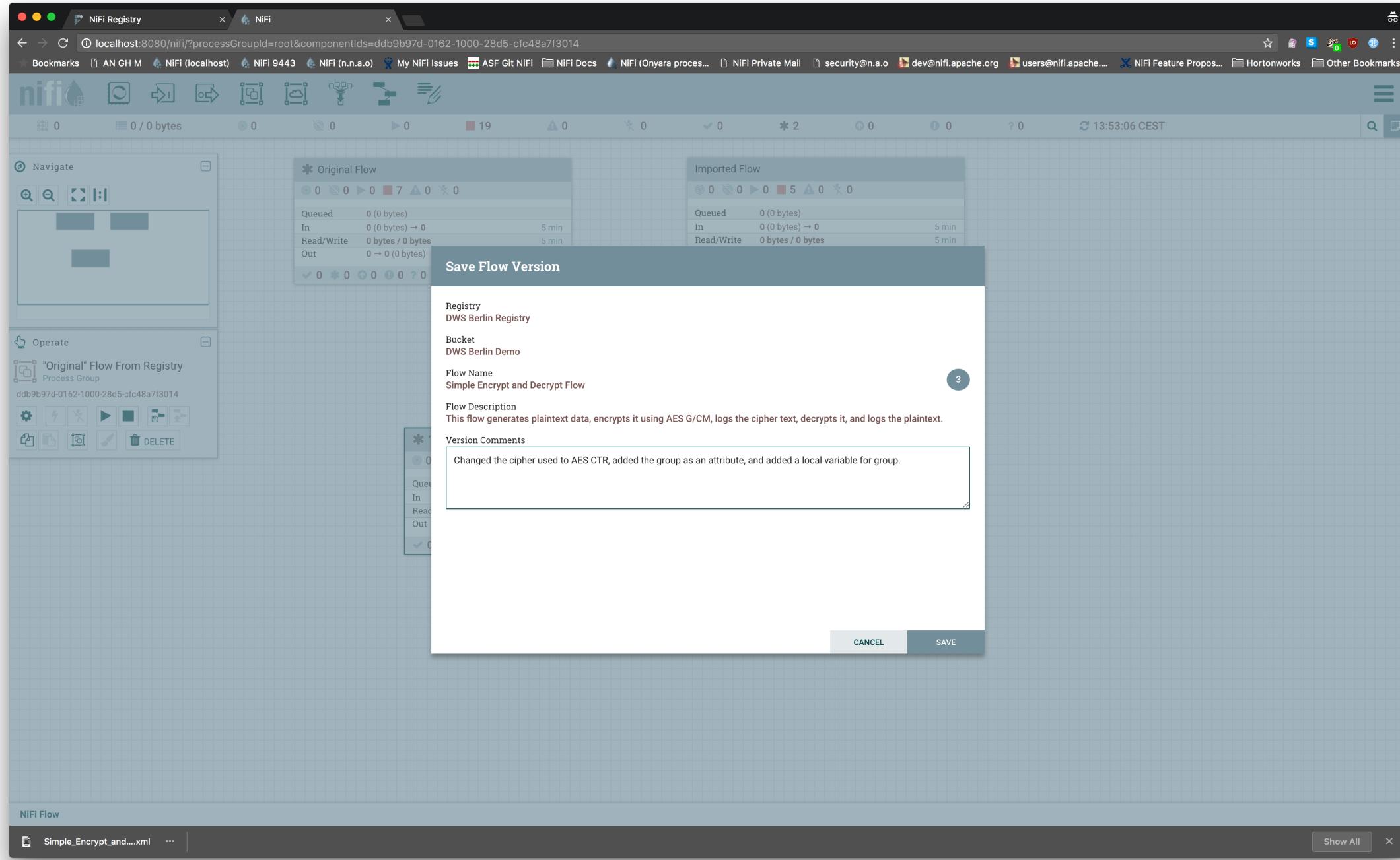
See local changes before committing



See local changes before committing



Commit



Commit

The screenshot shows a web browser window for the NiFi Registry at localhost:18080/nifi-registry/explorer/grid-list. The page displays a single flow entry:

Simple Encrypt And Decrypt Flow - DWS Berlin Demo (Flow)

VERSIONS: 2

BUCKET IDENTIFIER: d7df1fc1-d14b-4bb5-a8a6-ea57a60c5c98

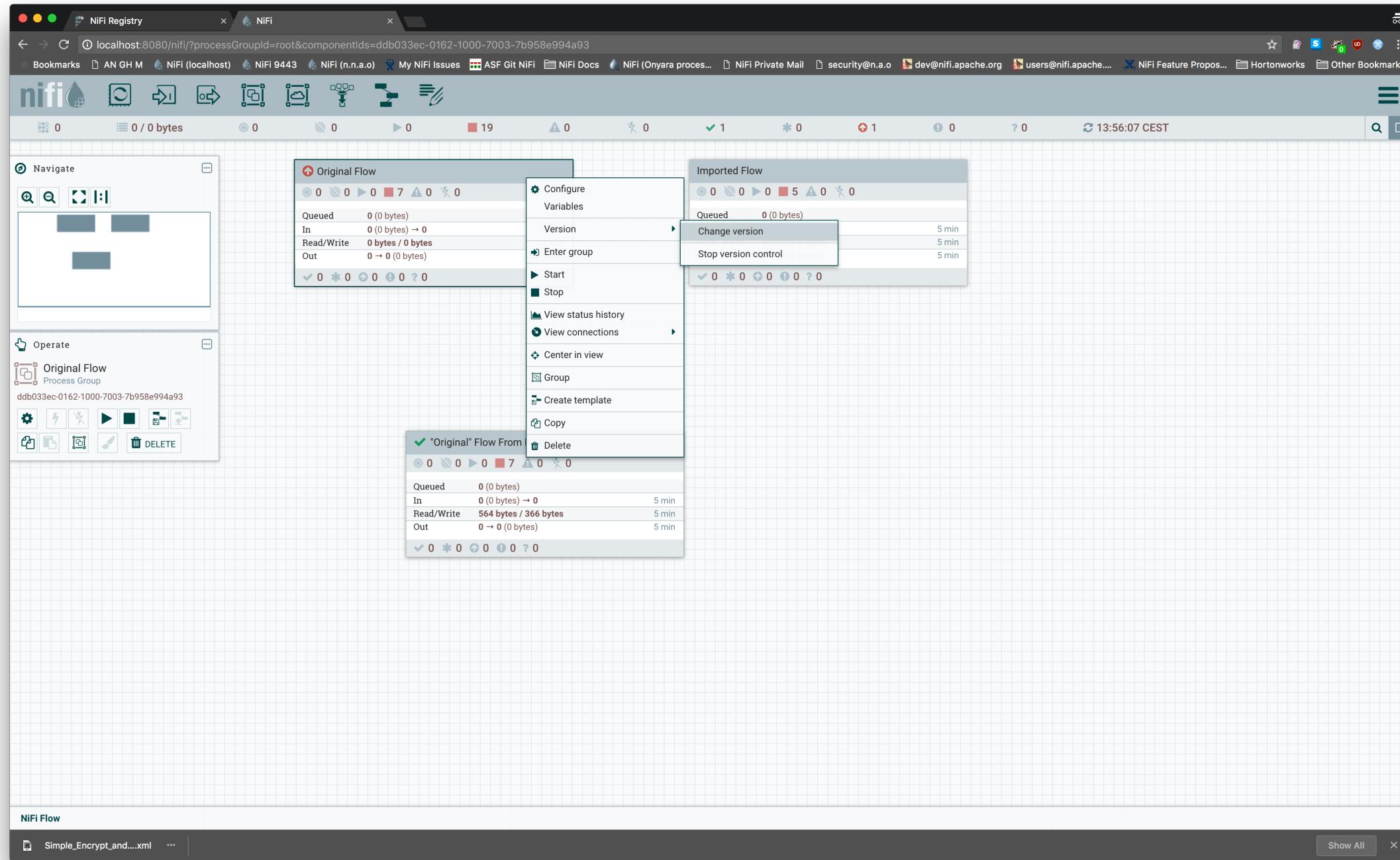
FLOW IDENTIFIER: b288f8eb-8adf-455a-b887-785239e97d33

DESCRIPTION: This flow generates plaintext data, encrypts it using AES GCM, logs the cipher text, decrypts it, and logs the plaintext.

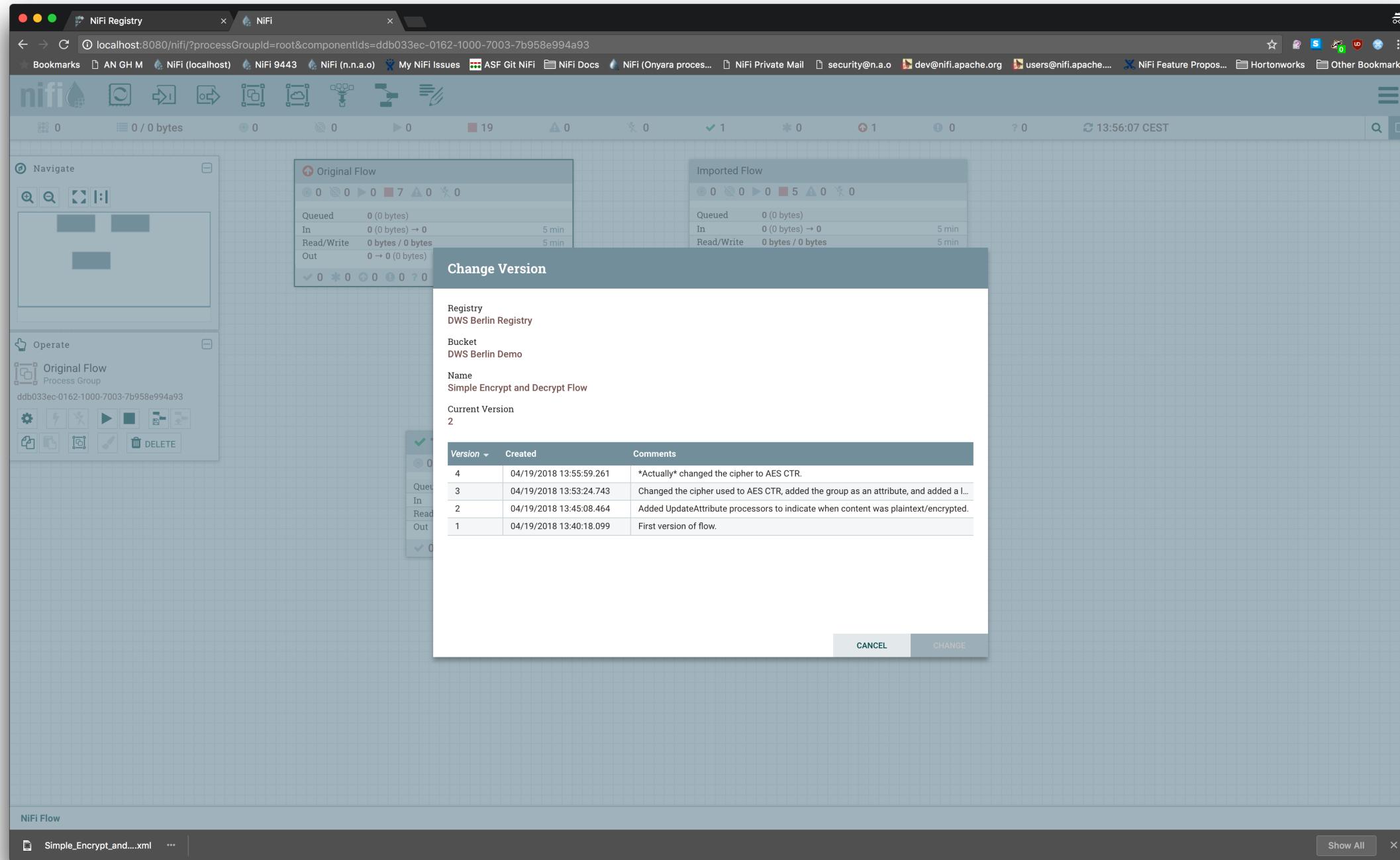
CHANGE LOG:

- Version 3 - a few seconds ago by anonymous: Changed the cipher used to AES CTR, added the group as an attribute, and added a local variable for group. (Apr-19-2018 at 1:53 PM)
- Version 2 - 8 minutes ago by anonymous
- Version 1 - 13 minutes ago by anonymous

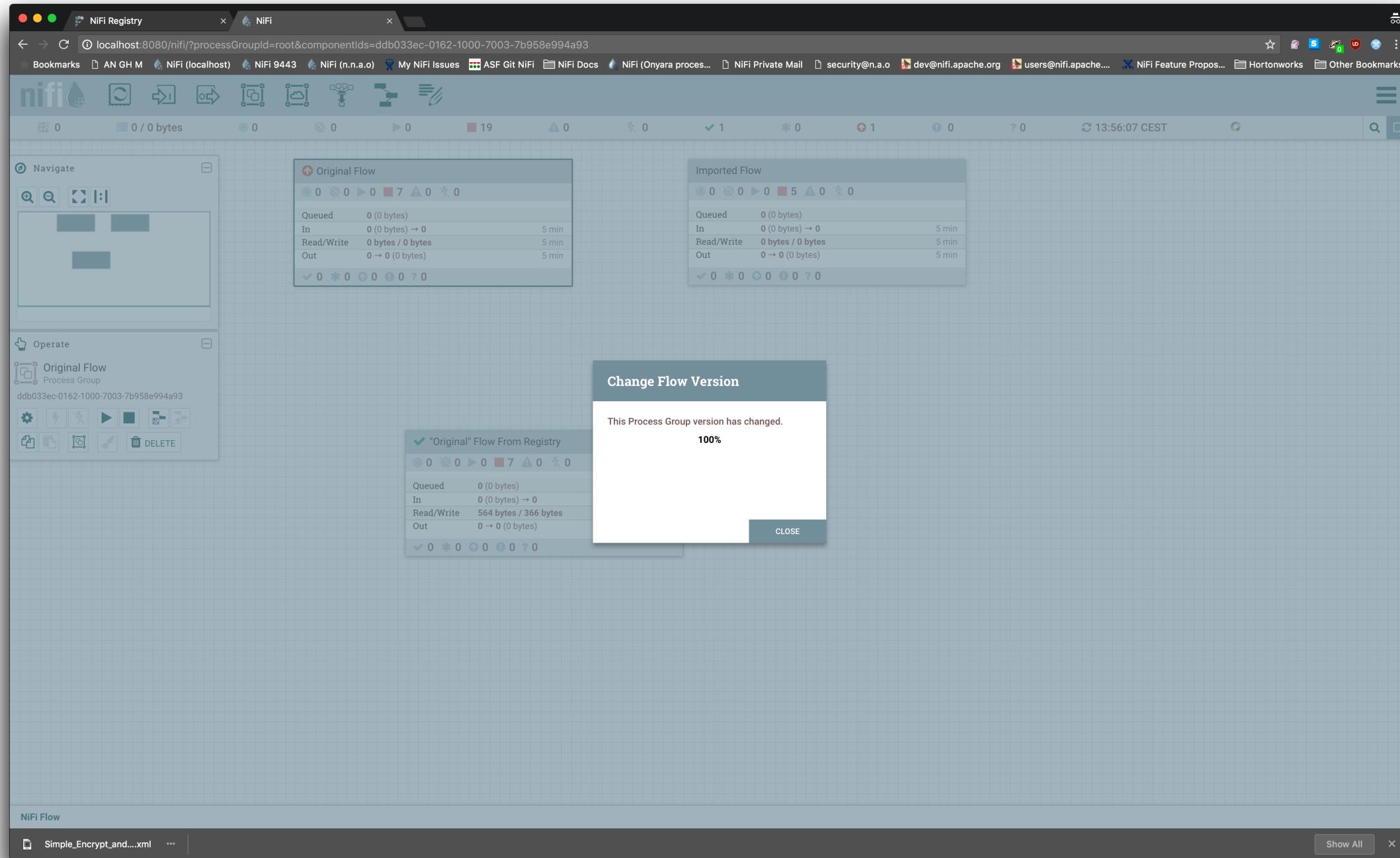
Update new instance from Registry



Update new instance from Registry



Update new instance from Registry



Future Direction and Likely Concerns

Other Features

- User access control
- Toolkit
- CLI
- Multiple (* - *) registries to NiFis

Key Takeaways

- Portability
 - Share with colleagues
 - Isolate responsibilities
- Versioning
 - Synchronize
 - Track changes
 - Ensure flow accuracy
 - Promotion between environments
- Collaboration
 - Reusable flow “portion” components

Future Enhancements

- git-backed persistence provider [NIFIREG-162 | PR 112]
- Commit hooks [NIFIREG-160 | PR 110]
- Preview incoming changes [NIFIREG-166]
- Your idea here?



Thank you

alopresto@hortonworks.com | alopresto@apache.org | [@yolopey](https://twitter.com/yolopey)
github.com/alopresto/slides

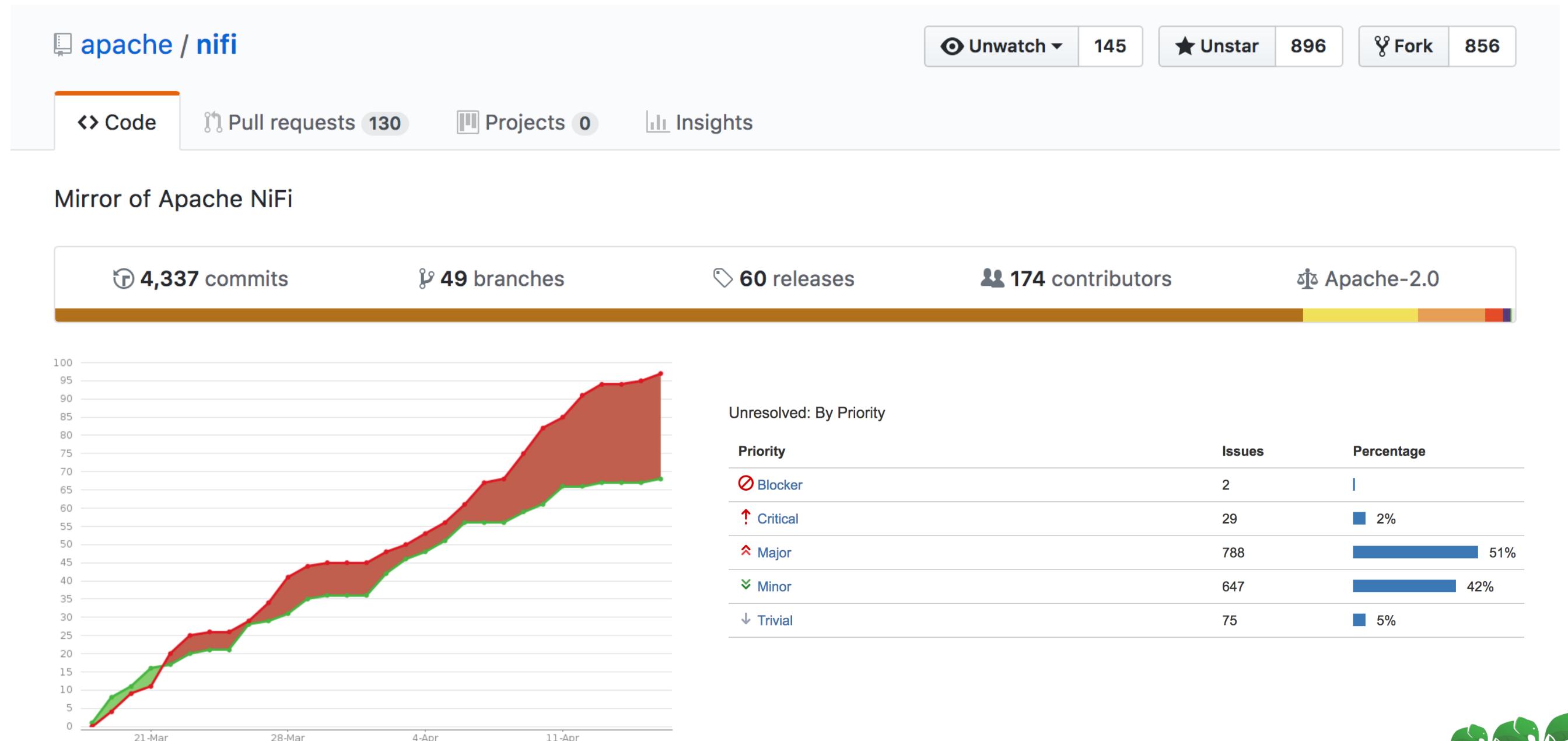
New Announcements

Introducing Apache NiFi Registry

- NiFi 1.6.0 – 08 April 2018
 - MongoDB, InfluxDB, Druid, HBase components
 - Granular @Restricted components
- MiNiFi C++ 0.4.0 – 27 January 2018
- MiNiFi Java 0.4.0 – 22 January 2018
- NiFi Registry 0.1.0 – 1 January 2018



Community Health



Learn more and join us

Apache NiFi site

<https://nifi.apache.org>

Subproject MiNiFi site

<https://nifi.apache.org/minifi/>

Subscribe to and collaborate at

dev@nifi.apache.org

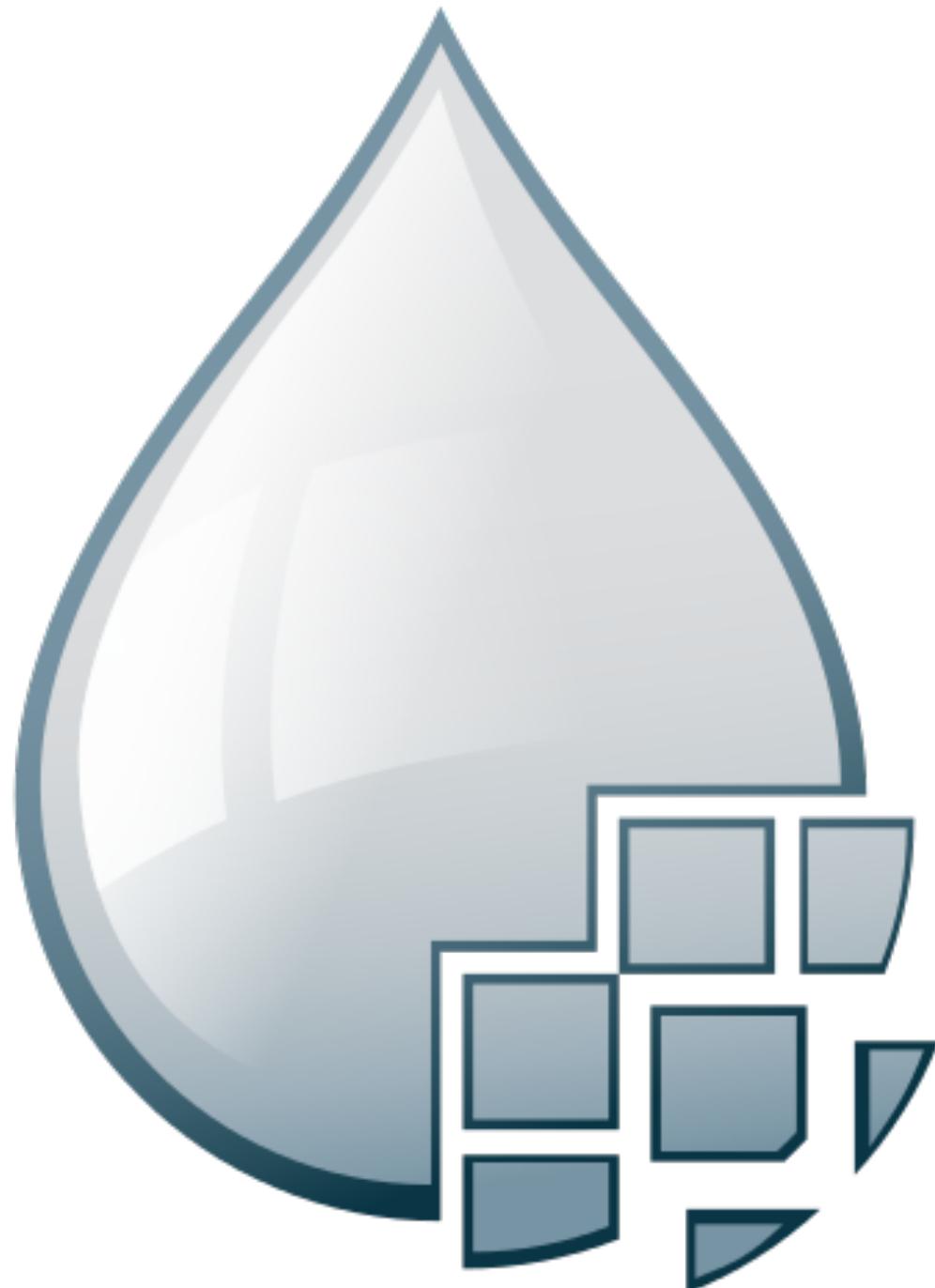
users@nifi.apache.org

Submit Ideas or Issues

<https://issues.apache.org/jira/browse/NIFI>

Follow us on Twitter

[@apachenifi](https://twitter.com/apachenifi)



More NiFi This Week...

Title	Room	Time	Speaker(s)
Apache NiFi Crash Course	Hall I - D	1115 - 1345	Andy LoPresto, Tim Spann
IoT with Apache MXNet and Apache NiFi and MiNiFi	Hall I - C	1150 - 1230	Tim Spann
Best practices and lessons learnt from Running Apache NiFi at Renault	Europe	1650 - 1730	Adel Gacem, Abdelkrim Hadjidj
From an experiment to a real production environment	Room V	1650 - 1730	Jeroen Wolffensperger, Martijn Groen
IoT, Streaming, and Dataflow Birds of a Feather	Room I	1740 - 1855	George Vetticaden, Davor Bonaci, Andy LoPresto, Stephan Ewen
Intelligently Collecting Data at the Edge — Intro to Apache MiNiFi	Room II	1100 - 1140	Andy LoPresto
The Power of Intelligent Flows: Realtime IoT Botnet Classification with Apache NiFi	Hall I - C	1400 - 1440	Andy LoPresto
Forget Duplicating Local Changes: Apache NiFi and the Flow Development Lifecycle (FDLC)	Room II	1600 - 1640	Andy LoPresto