

## Integration of Files with Alation Data Catalog

Information Asset has developed a solution to parse and transfer files of different file systems to Alation Data Catalog.

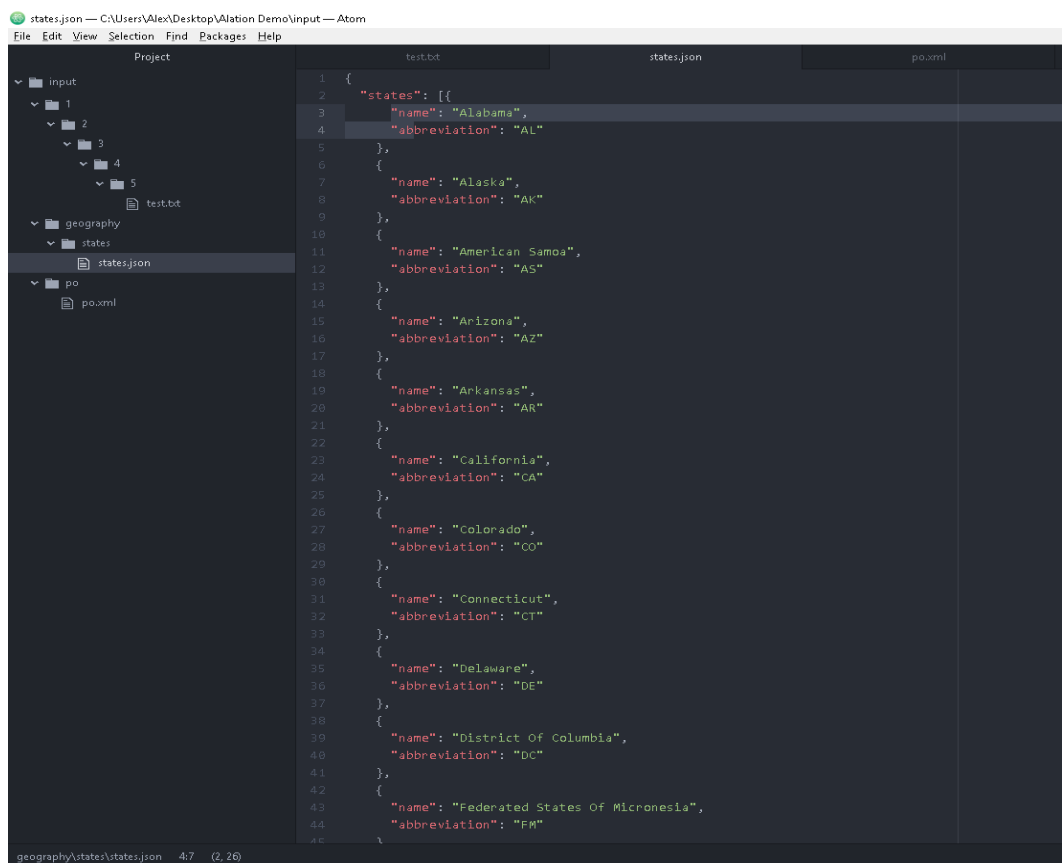
### Alation Data Catalog

Alation Data Catalog is a market-leading data catalog built for collaboration. Alation empowers analysts to find, understand and share data for faster, more accurate insights. Alation's data catalog automatically captures the rich context of enterprise data and keeps it updated with human-machine learning collaboration.

### Files

A file is used to store information. Without files, it would be difficult to understand where one piece of information stops and next begins. By separating the information and giving them names the information can be easily isolated and identified. There are different types of files such as JSON, XML, CSV, Parquet, and Avro so on.

Figure 1 shows the details of JSON file.



```
1 {
2   "states": [
3     {
4       "name": "Alabama",
5       "abbreviation": "AL"
6     },
7     {
8       "name": "Alaska",
9       "abbreviation": "AK"
10    },
11    {
12      "name": "American Samoa",
13      "abbreviation": "AS"
14    },
15    {
16      "name": "Arizona",
17      "abbreviation": "AZ"
18    },
19    {
20      "name": "Arkansas",
21      "abbreviation": "AR"
22    },
23    {
24      "name": "California",
25      "abbreviation": "CA"
26    },
27    {
28      "name": "Colorado",
29      "abbreviation": "CO"
30    },
31    {
32      "name": "Connecticut",
33      "abbreviation": "CT"
34    },
35    {
36      "name": "Delaware",
37      "abbreviation": "DE"
38    },
39    {
40      "name": "District Of Columbia",
41      "abbreviation": "DC"
42    },
43    {
44      "name": "Federated States Of Micronesia",
45      "abbreviation": "FM"
46    }
47  ]
48 }
```

Figure 1: JSON file.

Figure 2 shows a JSON file in Alation.

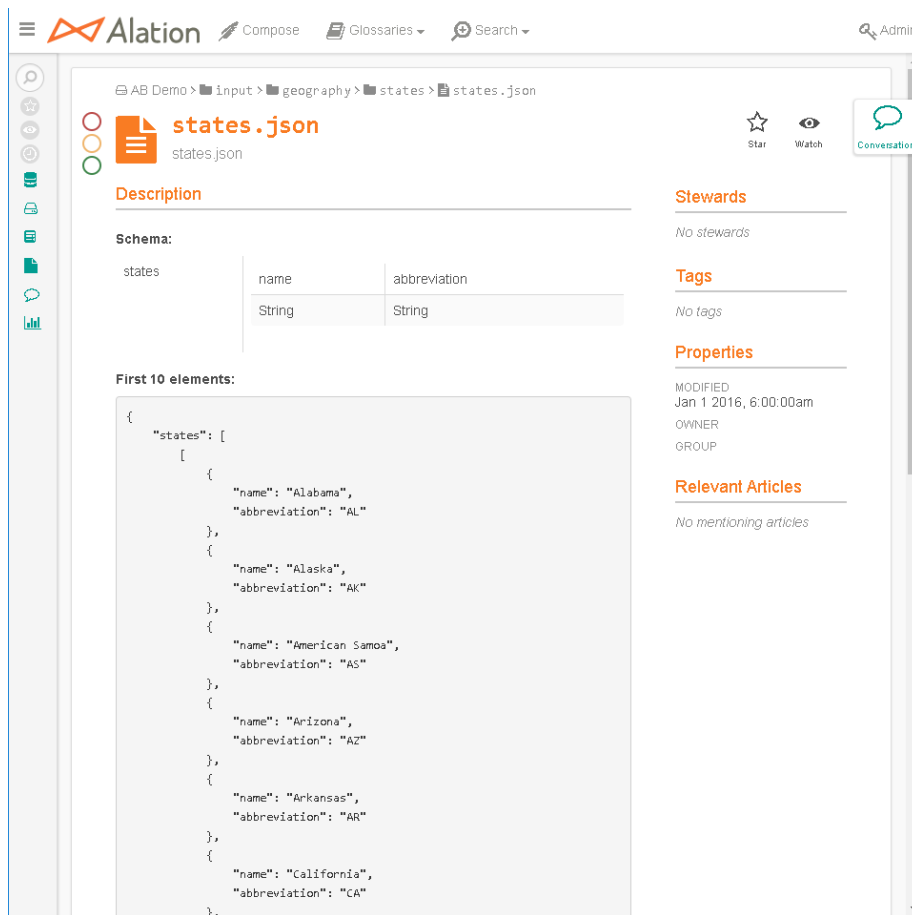


Figure 2: JSON file in Alation.

Figure 3 shows a CSV file.

Date rece	Product	Sub-prod	Issue	Sub-issue	Consumer	Company	Company	State	ZIP code	Tags	Consumer	Submitted	Date sent	Company	Timely re	Consumer	Complaint
#####	Mortgage	Other moi	Loan modification, collection, foreclosure		Dummy Comp	MI			177XX		N/A	Referral	03/17/201	Closed wi	Yes	No	759217
#####	Credit reporting	Incorrect i	Account s	I have out	Company	Dummy Comp	AL		352XX		Consent p	Web	#####	Closed wi	Yes	No	2141773
10/17/201	Consumer Vehicle lo	Managing the loan o	I purchased a new ca		Dummy Comp	PA			177XX	Older Am	Consent p	Web	10/20/201	Closed wi	Yes	No	2163100
#####	Credit card	Bankruptcy				Dummy Comp	ID		177XX	Older Am	N/A	Web	#####	Closed wi	Yes	Yes	885638
09/13/201	Debt colle	Credit cari	Communi	Frequent or repeated calls		Dummy Comp	VA		177XX		N/A	Web	09/13/201	Closed wi	Yes	Yes	1027760
11/13/201	Mortgage	Conventic	Loan servicing, payments, escrow account		Dummy Comp	MN			177XX		N/A	Phone	11/20/201	Closed wi	Yes	No	596562
06/16/201	Debt colle	Medical	Improper	Contacted employer	Company	Dummy Comp	CA		177XX		Consent n	Web	06/19/201	Closed wi	Yes	No	1422680
06/15/201	Credit reporting	Credit rep	Inadequat	An	Company	Dummy Comp	VA		177XX		Consent p	Web	06/15/201	Closed wi	Yes	No	1420702
11/13/201	Mortgage	Other moi	Loan modification, collection, fc	Company	Dummy Comp	CA			177XX		N/A	Referral	#####	Closed wi	Yes	Yes	1654890
10/21/201	Mortgage	Conventic	Loan modification, collection, foreclosure		Dummy Comp	FL			177XX	Older Am	N/A	Web	10/21/201	Closed wi	Yes	No	1079567
04/14/201	Mortgage	Other moi	Loan modification, collection, fc	Company	Dummy Comp	CA			177XX		N/A	Referral	04/14/201	Closed wi	Yes	No	2436165
03/30/201	Student lc	Non-fede	Repaying your loan			Dummy Comp	MN		177XX		N/A	Web	#####	Closed wi	Yes	No	42864
#####	Debt colle	Other	(i.e. Disclosure Not given	This company refuse		Dummy Comp	TX		177XX		Consent p	Web	#####	Closed wi	Yes	Yes	1772196
#####	Credit reporting	Incorrect i	Account status			Dummy Comp	NV		177XX		N/A	Referral	01/23/201	Closed wi	Yes	Yes	1184859
03/15/201	Credit card	Closing/Cancelling account				Dummy Comp	NY		177XX		N/A	Referral	03/18/201	Closed wi	Yes	No	357186
07/18/201	Credit reporting	Incorrect i	Account status			Dummy Comp	FL		177XX		N/A	Fax	07/19/201	Closed wi	Yes	No	2017178
02/17/201	Debt colle	Credit cari	Improper	Talked to This	Company	Dummy Comp	NE		177XX		Consent p	Web	#####	Closed wi	Yes	Yes	1790634
#####	Mortgage	Conventic	Loan modification, collection, foreclosure		Dummy Comp	MT			177XX		N/A	Web	#####	Closed wi	Yes	No	1107428
04/17/201	Mortgage	FHA mortg	Application, originator, mortgage broker		Dummy Comp	ME			177XX		Consent n	Web	04/17/201	Closed wi	Yes	No	1334421
#####	Credit reporting	Incorrect i	Information is not m	Company	Dummy Comp	CA			177XX		N/A	Phone	#####	Closed wi	Yes	No	1823789
#####	Debt colle	Mortgage	Cont'd att	Debt was paid		Dummy Comp	CA		177XX		N/A	Web	#####	Closed wi	Yes	Yes	1224106

Figure 3: CSV file.

Figure 4 shows the details of the CSV file in Alation.

LocalFiles>files>consumer-complai...

consumer-complaints.csv

files/consumer-complaints.csv

Star

Watch

Summary

No summary

Description

Edit

Schema:

Column	Type
Date received	String
Product	String
Sub-product	String
Issue	String
Sub-issue	String
Consumer complaint narrative	String
Company public response	String
Company	String
State	String
ZIP code	String
Tags	String
Consumer consent provided?	String
Submitted via	String
Date sent to company	String

Stewards

No stewards

Tags

No tags

Properties

MODIFIED  
Jun 11 2018, 2:59:26pm  
OWNER  
GROUP

Relevant Articles

No mentioning articles

Represented By

No represented by

Figure 4: CSV file in Alation.

Figure 5 shows the XML file.



```

1  <?xml version="1.0"?>
2  <PurchaseOrder PurchaseOrderNumber="99503" OrderDate="1999-10-20">
3    <Address Type="Shipping">
4      <Name>Ellen Adams</Name>
5      <Street>123 Maple Street</Street>
6      <City>Mill Valley</City>
7      <State>CA</State>
8      <Zip>10999</Zip>
9      <Country>USA</Country>
10   </Address>
11   <Address Type="Billing">
12     <Name>Tai Yee</Name>
13     <Street>8 Oak Avenue</Street>
14     <City>Old Town</City>
15     <State>PA</State>
16     <Zip>95819</Zip>
17     <Country>USA</Country>
18   </Address>
19   <DeliveryNotes>Please leave packages in shed by driveway.</DeliveryNotes>
20   <Items>
21     <Item PartNumber="872-AA">
22       <ProductName>Lawnmower</ProductName>
23       <Quantity>1</Quantity>
24       <USPrice>148.95</USPrice>
25       <Comment>Confirm this is electric</Comment>
26     </Item>
27     <Item PartNumber="926-AA">
28       <ProductName>Baby Monitor</ProductName>
29       <Quantity>2</Quantity>
30       <USPrice>39.98</USPrice>
31       <ShipDate>1999-05-21</ShipDate>
32     </Item>
33   </Items>
34 </PurchaseOrder>
35

```

Figure 5: XML file.

Figure 6 shows the details of an XML file in Alation.

LocalFiles

files

po.xml

po.xml

files/po.xml

Star

Watch

Summary

No summary

Description

Schema:

Field	Type
PurchaseOrderNumber	String
OrderDate	String
Address	Object
Address.Type	String
Address.Name	String
Address.Street	String
Address.City	String
Address.State	String
Address.Zip	String
Address.Country	String
DeliveryNotes	String
Items	Object
Items.Item	Object

Stewards

No stewards

Tags

No tags

Properties

MODIFIED

Jun 11 2018, 2:59:26pm

OWNER

GROUP

Relevant Articles

No mentioning articles

Represented By

No represented by

Figure 6: XML file in Alation.

Figure 7 shows the Avro file.



Figure 7: Avro file.

Figure 8 shows the details of an Avro file in Alation.

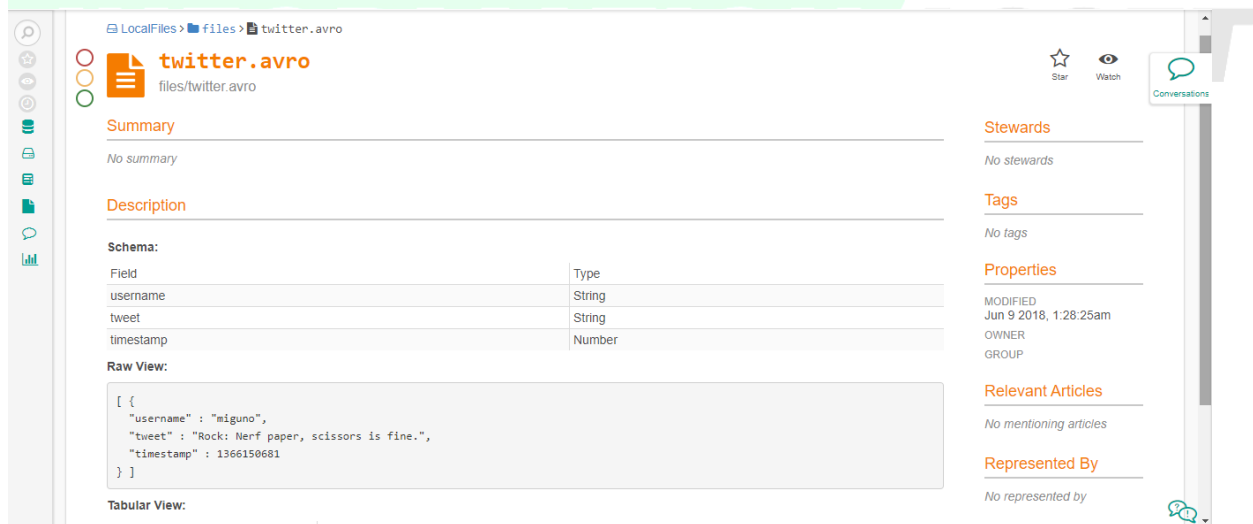


Figure 8: Avro file in Alation.

Figure 9 shows the Parquet file.

registration_dttm	id	first_name	last_name	email	gender	ip_address	cc	country	birthdate	salary	title
03-02-2016 13:36:39	1	Donald	Lewis	dlewis0@clt	Male	102.22.124.20		Indonesia	7/9/1972	140249.37	Senior Finc
03-02-2016 00:22:28	2	Walter	Collins	wcollins1@b	Male	247.28.26.93	3587726	China			
03-02-2016 18:29:04	3	Michelle	Henderson	mhenderson	Female	193.68.146.150		France	1/15/1964	236219.26	Teacher
03-02-2016 13:42:19	4	Lori	Hudson	lhudson3@		34.252.168.48	3568840	Russia	4/22/1988		Nuclear Pc
03-02-2016 00:15:29	5	Howard	Miller	hmiller4@fer	Male	103.193.150.230	3583473	France	11/26/1998	50210.02	Senior Edit
03-02-2016 10:49:07	6	Frances	Adams	fadams5@1	Female	106.196.106.93		Russia	3/27/1997	82175.77	Account C
03-02-2016 19:44:12		Steven	Hanson	shanson6@c	Male	234.130.172.185	3550842	Indonesia		129582.61	
03-02-2016 08:11:34	8	Louis	Simmons	lsimmons7@	Male	18.69.80.15		China	6/1/1992	90744.86	Product Er
03-02-2016 23:56:51	9	Keith	Parker	kparker8@a	Male	108.205.40.64		Guadeloupe	12/30/1992	60618.9	Developer
03-02-2016 07:47:29	10	Wanda	Walker	wwalker9@k	Female	246.214.98.78	3539421	Portugal		137664.53	
03-02-2016 16:39:23	11	Kathryn	Weaver	kweavera@	Female	157.237.161.75	2014250	Sweden		117572.65	
03-02-2016 06:50:56	12	Philip	Ward	pwardb@sa	Male	77.140.225.69	2015080	Greece	9/3/1984	238925.79	Human Re
03-02-2016 23:36:58	13	Evelyn	Harvey	eharveyc@t		254.174.154.7	3539535	China	5/15/1979		Software E
03-02-2016 07:33:24	14	Andrea	Lane	alaned@go	Female	192.253.116.192	5100174	Indonesia	1/19/1989	166778.42	Operator
03-02-2016 13:45:04	15	Bobby	Vasquez	bvasqueze@	Male	126.60.18.195	3581051	Philippines	1/25/1975	138184.83	Senior Edit
03-02-2016 21:58:10	16	Kenneth	Gibson	kgibsonf@sc	Male	91.153.142.170	5389947	Peru	11/3/1975	98614.53	Environme
03-02-2016 16:22:34	17	Emily	Hill	ehillg@hous	Female	109.107.174.205		Palestinian Ter	5/18/1956	218781.48	Executive
03-02-2016 23:20:49	18	Kelly	Fowler	kfowlerh@d	Female	147.58.88.116	3551741	Greece	6/11/1975	117249.56	Systems Ac
03-02-2016 18:28:46	19	Diana	Howell	dhowell@sc	Female	21.240.75.42	4026635	Iran	7/7/1993	174844.52	Teacher
03-02-2016 02:23:26	20	Johnny	Collins	jcollinsj@goc	Male	38.173.129.250	3723016	Afghanistan	7/28/1987	155908.69	Social Wor
03-02-2016 10:05:12	21	Frank	Bradley	fbradleyk@s	Male	186.9.38.46	4913033	Czech Republ		211051.83	

Figure 9: Parquet file.

Figure 10 shows the details of Parquet file in Alation.

The screenshot shows the Alation web interface. The top navigation bar includes the Alation logo, 'Compose', 'Glossaries', and a search bar. On the right, there are links for 'Admin', 'Help', and a user profile. The main content area displays the details for a file named 'userdata2.parquet' located at 'files/userdata2.parquet'. The file is represented by an orange document icon. Below the file name, there are sections for 'Summary' (showing 'No summary'), 'Description', and 'Schema'. The 'Schema' section is a table with two columns: 'Field' and 'Type'. The fields listed are: cc (String), country (String), birthdate (String), comments (String), gender (String), registration\_dttm (String), last\_name (String), ip\_address (String), salary (String), and title (String). To the right of the schema, there are sections for 'Stewards' (No stewards), 'Tags' (No tags), 'Properties' (MODIFIED: Jun 9 2018, 1:28:25am; OWNER; GROUP), 'Relevant Articles' (No mentioning articles), and 'Represented By' (No represented by). The interface also includes a sidebar with various icons and a 'Conversations' button on the right.

Figure 10: Parquet file in Alation.

## Integration of Files and Alation

Information Asset has developed a solution that enables a user to import JSON, CSV, XML, Avro, and Parquet file types into Alation Data Catalog. The data scientist can view the data structure along with sample data.

## Next Steps

To learn more about this solution, please request a demo by contacting [sales@information-asset.com](mailto:sales@information-asset.com) or visit our website at [www.information-asset.com](http://www.information-asset.com).

Copyright © 2018, Information Asset, LLC. All rights reserved.

This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor is it subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.



**INFORMATION ASSET**



**INFORMATION ASSET**