



Joint Organisations Data Initiative

JODI-Oil: Data Quality Assessment and Checks

Stève Gervais – Head of the Monthly Oil and Gas Statistics Unit - IEA

17th JODI Training Workshop – 11-12 December 2018, Beirut, Lebanon

- Data quality is a multi-dimensional concept
 - Relevance (of statistical concepts)
 - Accuracy
 - Timeliness
 - Accessibility and clarity of information
 - Comparability of statistics
 - Coherence
 - Completeness/coverage
 - Cost and burden
 - Metadata

- The relevance of statistical information reflects the degree to which the information meets the current and potential needs of key users.
- Identification of the users and their expectations is very important
- Consult with oil companies in the country
 - How many are they? How important is each of them?
 - Listen to their expectations and needs (synthesize and prioritize)
 - Convince them to follow definitions, methodology, classifications (if not possible, at least keep a record of discrepancies)
- Reasons for collecting specific data
- Relevance varies according to countries/regions



- Accuracy is defined as the closeness between the computations or estimates and the (unknown) true population value (benchmark???)
- Assessing the accuracy of an estimate involves analysing the total error associated with the estimate: Bias (+ or -?) and standard deviation
- Sampling errors: due to problems in the design of a sample survey
- Non-sampling errors:
 - Coverage errors
 - Measurement errors
 - Processing errors
 - Non-response errors

Usually negatively correlated to timeliness & completeness



Data quality checks

- Indicate units in which you submit the data
- Look carefully at conversion factors if used
- Check signs of the values
 - Negative calculated refinery intake or demand indicates a problem in the balance.
- Check if totals match sum of components

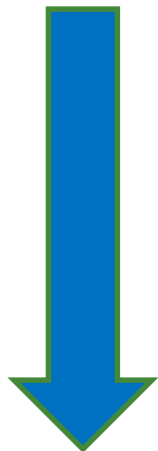
Balance check

- Primary (oil)
- Secondary (petroleum products)

Country _____

Month _____

Unit : _____



	Crude Oil	NGL	Other	Total (1)+(2)+(3)		Petroleum Products							Total Products (5)+(6)+(7) +(8)+(10) +(11)+(12)	
						LPG	Naphtha	Gasoline	Total Kerosene	Of which: Jet Kerosene	Gas/ Diesel Oil	Fuel Oil		Other Products
	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
+ Production					+ Refinery Output									
+ From Other sources					+ Receipts									
+ Imports					+ Imports									
- Exports					- Exports									
+ Products Transferred /Backflows					- Products Transferred									
- Direct Use					+ Interproduct Transfers									
- Stock Change					- Stock Change									
- Statistical Difference	0	0	0	0	- Statistical Difference	0	0	0	0	0	0	0	0	0
= Refinery Intake					= Demand									
Closing stocks					Closing stocks									

Balance check

- High statistical differences can be a red flag (more than 5-10%).
- A statistical difference of 0 is also suspicious!

	Crude Oil	NGL	Other	Total (1)+(2)+(3)		Petroleum Products								
						LPG	Naphtha	Gasoline	Total Kerosene	Of which: Jet Kerosene	Gas/ Diesel Oil	Fuel Oil	Other Products	Total Products (5)+(6)+(7) +(8)+(10) +(11)+(12)
	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
+ Production					+ Refinery Output									
+ From Other sources					+ Receipts									
+ Imports					+ Imports									
- Exports					- Exports									
+ Products Transferred /Backflows					- Products Transferred									
- Direct Use					+ Interproduct Transfers									
Stock Change					- Stock Change									
- Statistical Difference	0	0	0	0	- Statistical Difference	0	0	0	0	0	0	0	0	0
= Refinery Intake					= Demand									
Closing stocks					Closing stocks									

- Calculated refinery intake \approx reported refinery intake
- Calculated refinery intake = production + from other sources + imports – exports + products transferred/backflows – direct use – stock change
- Calculated refinery intake - reported refinery intake (statistical difference) should ideally be relatively small
- Statistical difference should be small in relative and absolute terms

Balance check – primary oil

		Crude oil (kt)
+	Production	2
+	From other sources	0
+	Imports	3681
-	Exports	0
+	Products transferred/backflows	0
-	Direct use	200
-	Stock change	-295
-	Statistical difference: calculated – reported refinery intake	228
=	Refinery intake	3550
%	Percentage statistical difference	6.4%

- Calculated demand \approx reported demand
- Calculated demand = refinery output + receipts + imports – exports - products transferred + interproduct transfers – stock change
- Calculated demand - reported demand should ideally be relatively small
- Statistical difference should be small in relative and absolute terms

Balance check – secondary oil products

		Total products (kt)
+	Refinery output	126
+	Receipts	0
+	Imports	59
-	Exports	13
-	Products transferred	0
+	Interproduct transfers	0
-	Stock change	-2
-	Statistical difference: calculated – reported demand	-2
=	Demand	176
%	Percentage statistical difference	-1%

- Applicable only if all data are complete and reliable
- Large deviations would require review and/or verification/correction
- Re-submission in the form of revisions during the following month
- Application on every column of the JODI oil questionnaire
- Range of 5% quite large for physical quantities (1% or even 0.5%)

Internal consistency check

- Another indicator of accuracy
- Fuel checks – total oil products should be equal to the sum of reported products (excluding jet fuel)
- Statisticians should ensure that this property holds in all submissions

Country _____

Month _____



Unit : _____

	Crude Oil	NGL	Other	Total (1)+(2)+(3)		Petroleum Products							Total Products (5)+(6)+(7) +(8)+(10) +(11)+(12)	
						LPG	Naphtha	Gasoline	Total Kerosene	Of which: Jet Kerosene	Gas/ Diesel Oil	Fuel Oil		Other Products
	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
+ Production					+ Refinery Output									
+ From Other sources					+ Receipts									
+ Imports					+ Imports									
- Exports					- Exports									
+ Products Transferred / Backflows					- Products Transferred									
- Direct Use					+ Interproduct Transfers									
- Stock Change					- Stock Change									
- Statistical Difference	0	0	0	0	- Statistical Difference	0	0	0	0	0	0	0	0	0
= Refinery Intake					= Demand									
Closing stocks					Closing stocks									

Internal consistency check

JOINT OIL DATA INITIATIVE

Closing minus opening level
Positive number corresponds to stock build, negative number corresponds to stock draw

Country Country

Month Month Year

Unit : thousand tons

	Crude Oil	NGL	Other	Total (1)+(2)+(3)	Petroleum Products								Total Products (5)+(6)+(7)+ (8)+(10)+ (11)+(12)	Checks
					LPG	Naphtha	Gasoline	Total Kerosene	Of which: Jet Kerosene	Gas/ Diesel Oil	Fuel Oil	Other Products		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
+ Production	12622	1883	3954	18,459	+ Refinery Output	125	274	2559	517	455	2536	397	1147	7,555
+ From Other sources			0	0	+ Receipts	0	108	622	13	10	125	36	1487	2,391
+ Imports	2453	59	0	2,512	+ Imports	6	0	229	156	127	86	90	393	960
- Exports	9066	969	2310	12,345	- Exports	53	54	605	43	43	695	243	202	1,895
+ Products Transferred /Backflows			536	536	- Products Transferred	0	25	0	0	0	0	2	509	536
- Direct Use	0	602	0	602	+ Interproduct Transfers	216	-18	169	-23	-10	105	-26	-423	0
- Stock Change	1012	315	0	1,327	- Stock Change	28	-50	-63	-33	-44	16	39	-87	-150
- Statistical Difference	911	43	0	954	- Statistical Difference	46	30	49	11	12	76	53	87	108
= Refinery Intake	5908	99	2180	8,187	= Demand	312	305	2988	642	571	2217	160	1893	8,517
Closing stocks	9246	1973	0	11,219	Closing stocks	258	100	1712	338	306	1757	315	1253	5,733

Automatic Checks

Total sum	OK
Statistical Difference	OK
Stat. Diff./Refinery Intake	Statistical Difference above 10% of Refinery Intake, please investigate
Products Transferred	OK
Negative Products Transferred	OK
Blocked out cells	OK
Negative Stock Values	OK
Refinery Losses	632 OK

Automatic Checks Petroleum Products

Total Products sum	OK
Statistical Difference	OK
Stat. Diff./Demand	Statistical Difference above 10% of Demand, please investigate
Negative Products Transferred	OK
Interproduct transfers	OK
Jet Kerosene	OK
Negative Stock Values	OK

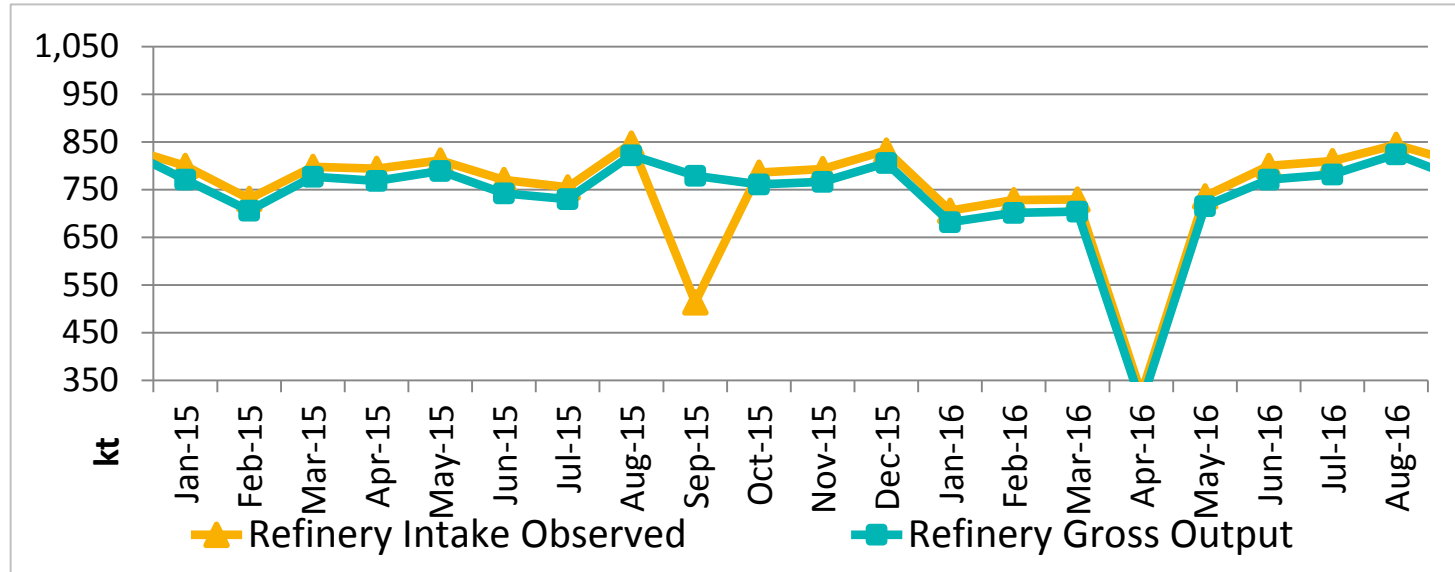
- Stock checks
- Stock change (M) = Closing stock level (M) – Closing stock level (M-1)
- Calculated stock change \approx reported stock change

Gas/Diesel Oil	Oct-16	Nov-16
+ Refinery Output	317.295	264.679
+ Receipts	0	0
+ Imports	0	0
- Exports	204.108	248.903
- Products transferred	0	0
+ Interproduct transfers	0	0
- Stock change	82.281	-126.009
- <i>Statistical difference</i>	0	0
= Demand	30.906	141.785
Closing stocks	289.78	274.286

Calculated stock change = closing stock level in Nov 2016 - closing stocks level in Oct 2016 **-15.494**

- Similar percentage changes (m-o-m, y-o-y) indicator of “good” data quality
- “Unusually large” percentage changes may require verification of data
- Seasonality of oil data
- Trade data
- Refinery intake/output check
 - Refinery yield (%) = Refinery output (total oil) products)/Refinery intake (total primary products) *100

Time series check – example 1



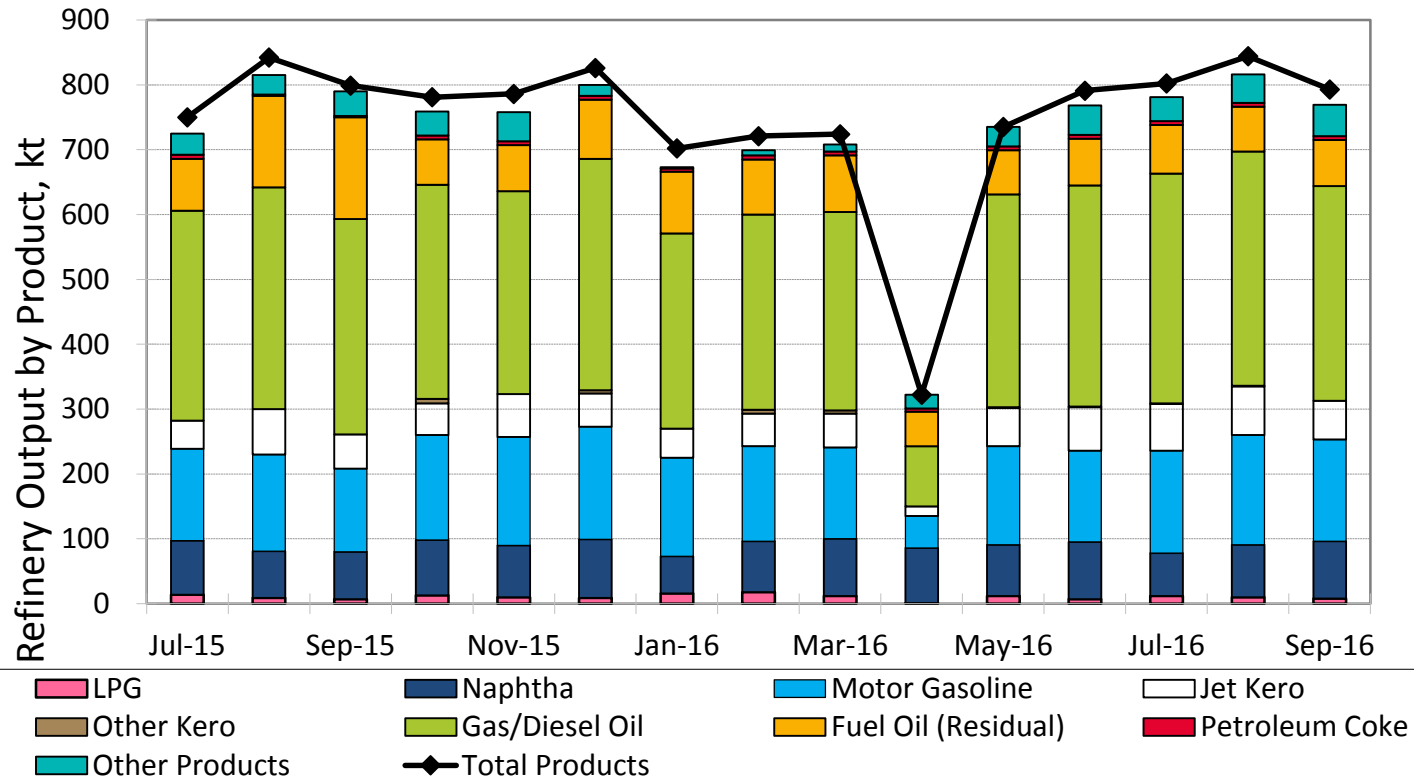
Refinery Output > Refinery Intake = Refinery Gains

Refinery Output < Refinery Intake = Refinery Losses

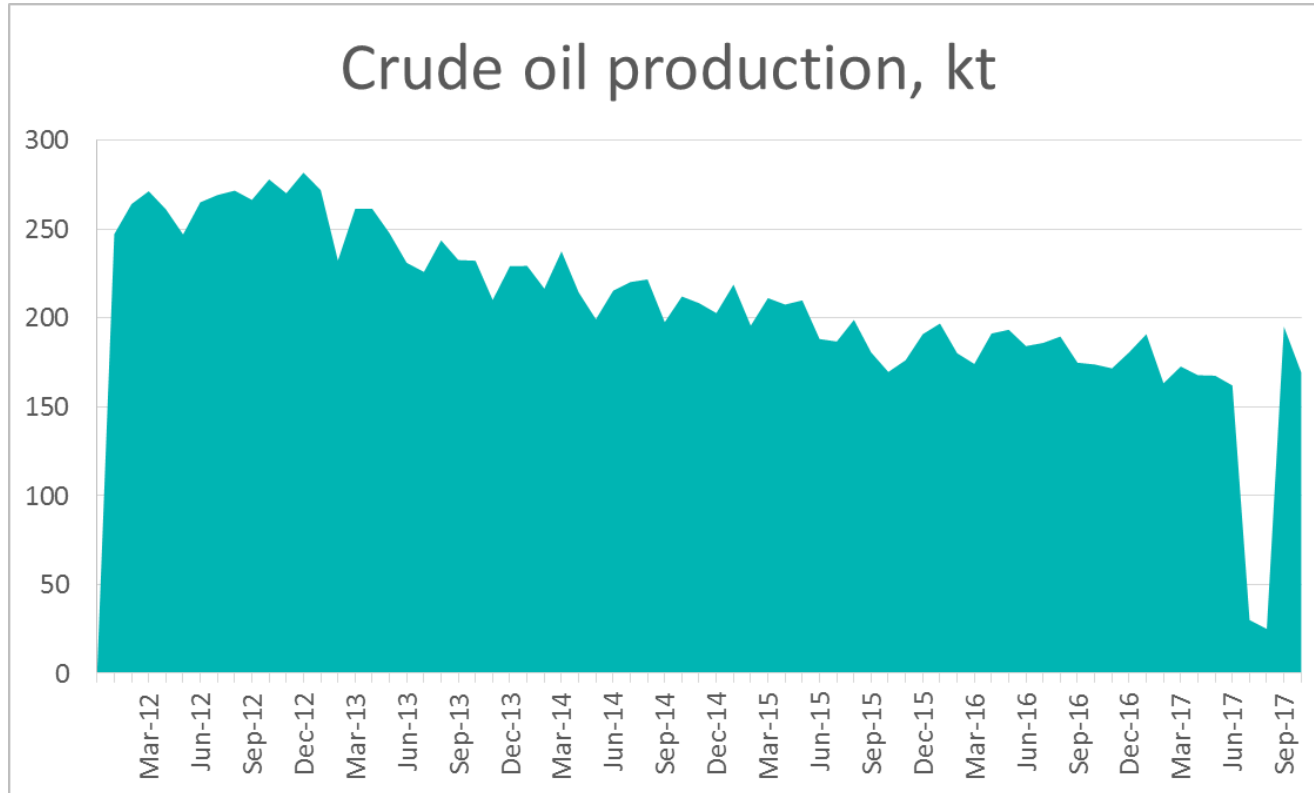
The units matter:

- Mass units (tons) – refinery output should be smaller than refinery intake but close to 100% with no gains.
- Volume units (bbls, litres) – either gains or losses are possible.

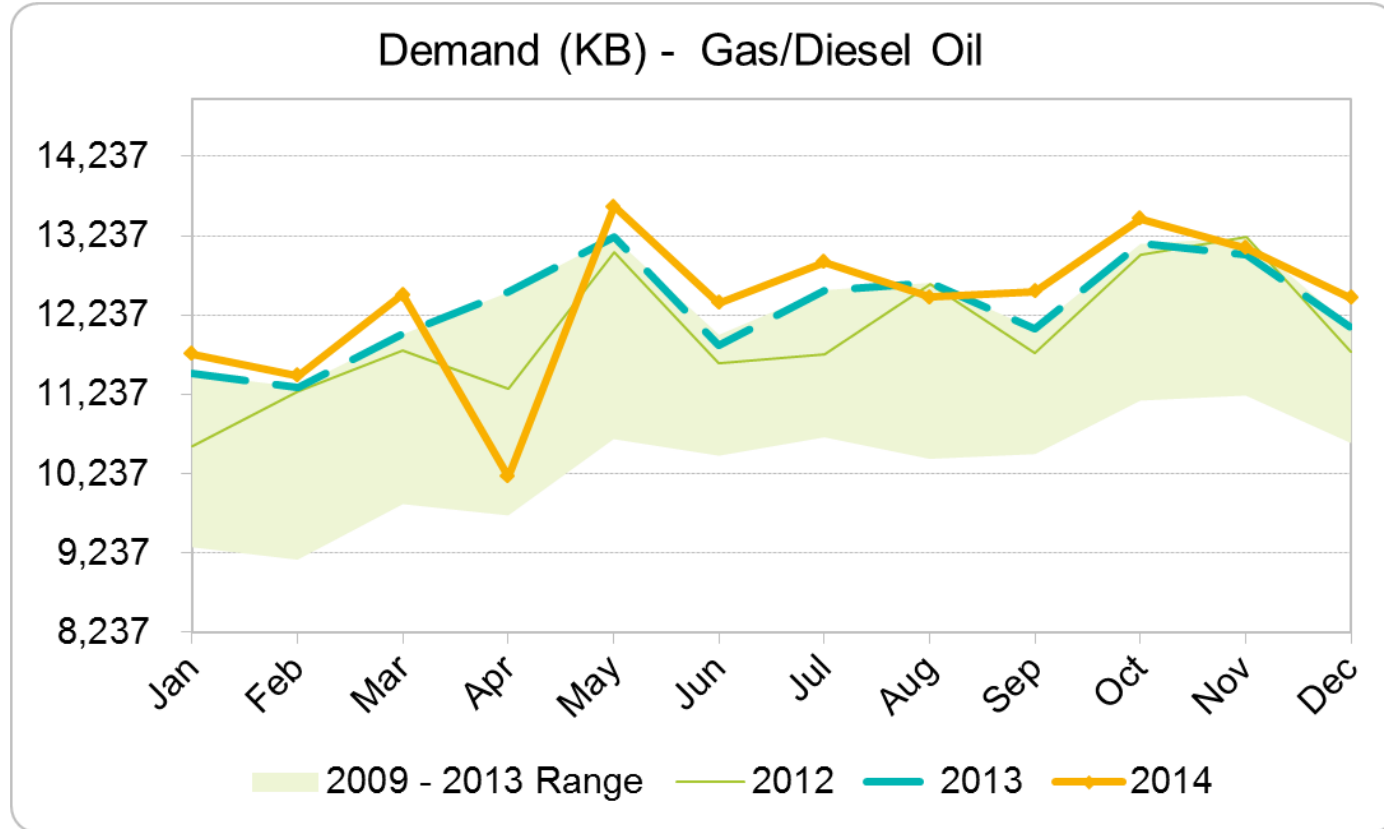
Time series check – example 2



Time series check – example 3



Time series check – example 4



- Do we cover all the market in monthly data collection?
- If we cover it only partially, what share do we cover?
- Should we estimate the missing part?
- Does the part which we cover shrink over time?

- Unified definitions/comprehensive metadata
- Knowledge of the conversion factors at country level
- Common unit of measurement
- Transparent methodology
- Timely submission of data

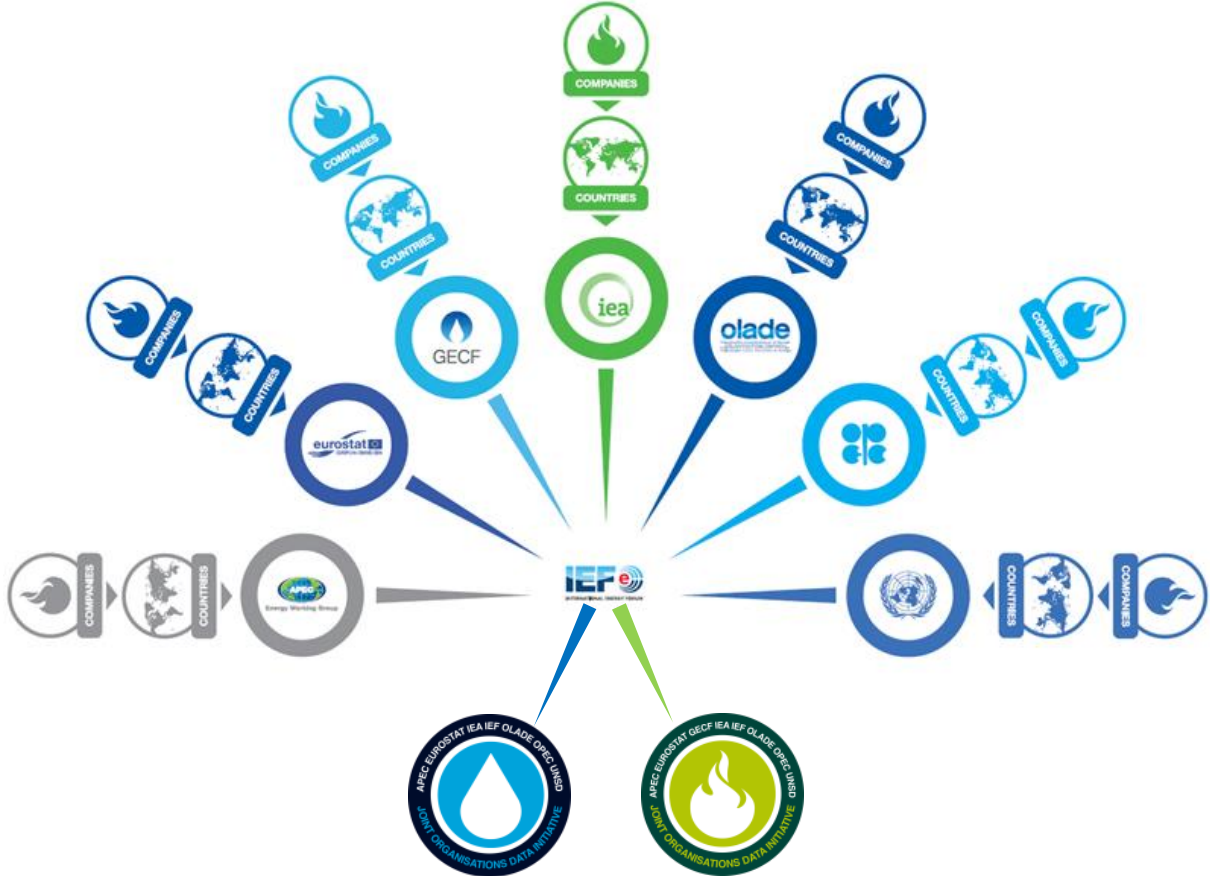
- The quality of the data will be affected by available resources to collect, analyze and store energy statistics
- Although not measures of quality, they are positively correlated with quality
- Costs: Specialized equipment, office space, utility bills, staff-hours involved, software, etc.
- Response burden: Simplest way to measure is the time spent by the respondent to provide information
- A compromise between quality and cost and burden must be achieved

- The simplest definition of metadata is that it is data about data. More specifically information (data) about a particular content (data)
- Metadata describes how and when and by whom a particular set of data was collected; how the data is formatted
- Metadata must be updated when there is a change in resource it describes
- It can be useful to keep metadata even when the resource no longer exists
- Metadata enhances data transparency and is essential for understanding information stored in a database

- Extensive coverage
 - JODI Oil
 - 110+ countries
 - Around 90% of global supply and demand
- Official data
 - Data published is as submitted (after data quality exchanges)
 - Metadata included
 - No estimation for missing information (at organizational level)
- Centralized/harmonized dissemination
 - Published time table of updates
 - Beyond 20/20 and CSV format

JODI are short-term data

How JODI works



JODI World databases

- Company/Industry level
 - Collection of data from different production units
 - Submission of data to central administration
- Administration level
 - Evaluate validity of submitted data
 - Add other data (e.g. administrative data)
 - Estimate missing/incomplete information
 - Aggregate at a national level
 - Submit to international organization

- Interaction with administrations
 - Provide support for technical and methodological issues
 - Encourage complete submissions
 - Encourage revisions when data become final
 - Training and capacity building
- Ensuring the validity of the data
 - Continuous checking of data: balance, outliers etc.
 - Comparison with external data when available
 - Internal consistency checks for complete submissions
 - Problems detected are discussed with administrations
- Send data to IEF for final validation and publication

Country Level

Less confidentiality

More coverage

Higher quality data

Better timeliness

International Level

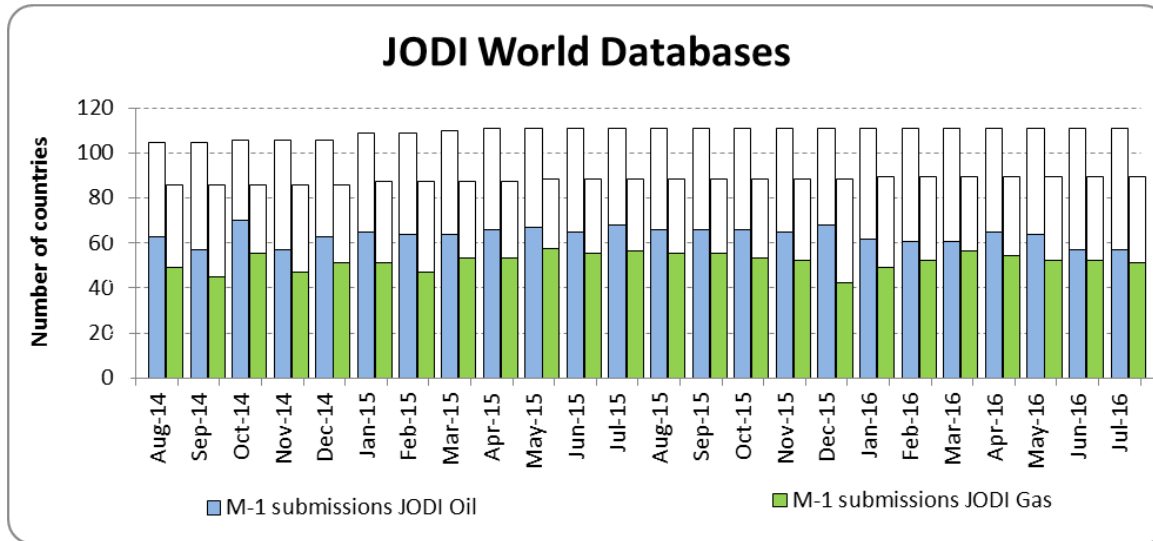
Harmonization

JODI World Database

Trainings

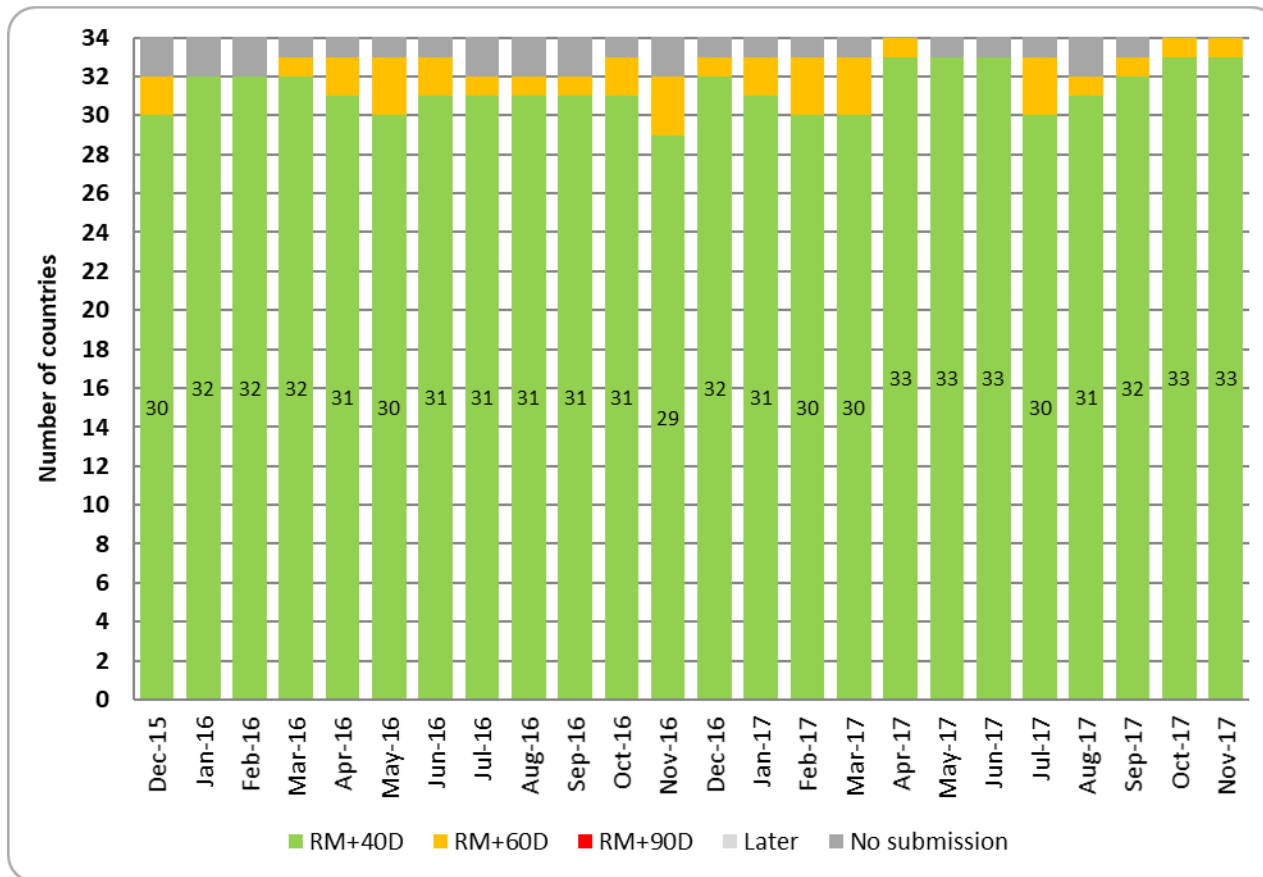
- Timeliness and Sustainability
- Completeness and Confidentiality
- Timeliness vs. Data Quality vs. Completeness

- Challenging M-1 deadline
 - 25th of the month following reference month
 - Multiple departments/agencies involved in collection
 - Time needed for checking and processing

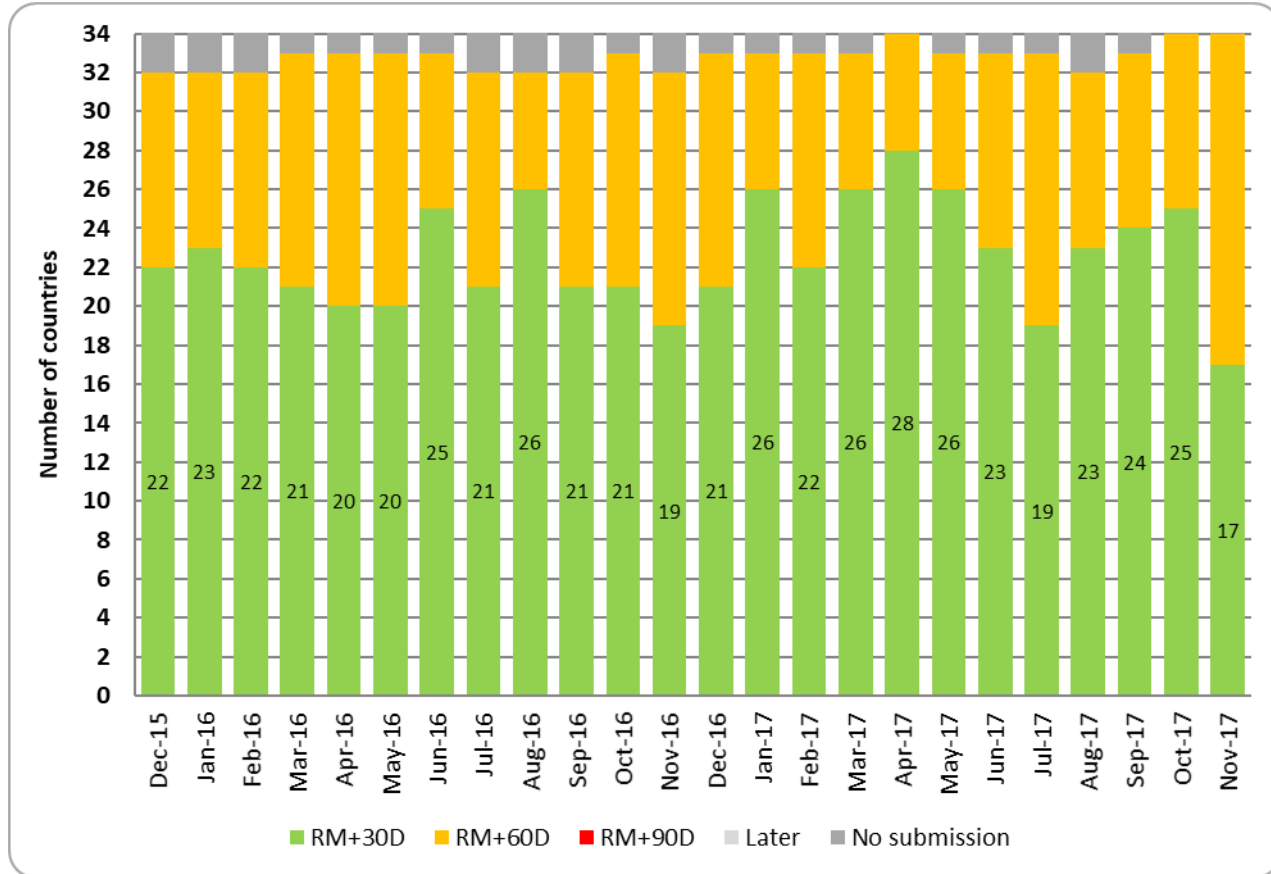


Difficult for all countries to meet the monthly deadline

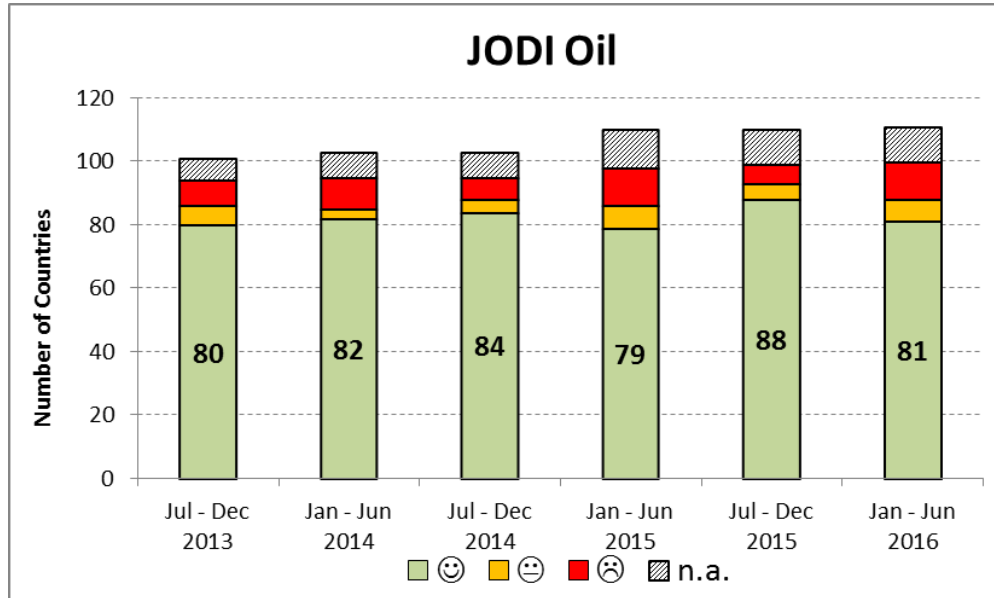
JODI Oil – Participation and Timeliness - IEA/OECD



JODI Oil – Participation and Timeliness (M-1=RM+30D) - IEA/OECD

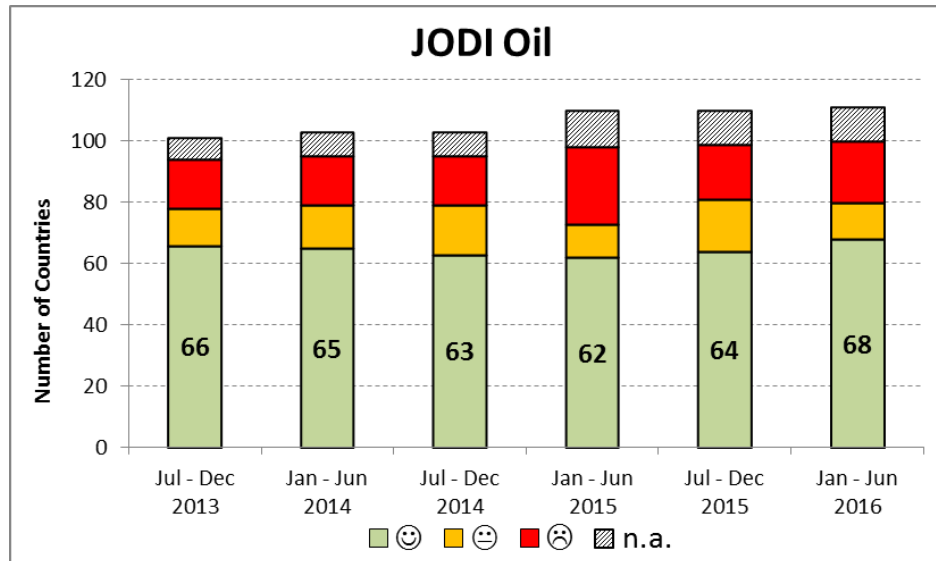


- To be of most use JODI needs regular monthly reporting
 - Some countries miss certain months
 - Some countries submit later (quarterly or annually)



Completeness and Confidentiality

- Availability of monthly data in countries
- Submission of revisions with more complete data
- Issues of confidentiality
 - Legislative issues/authorisation
 - Perceived as commercially sensitive information



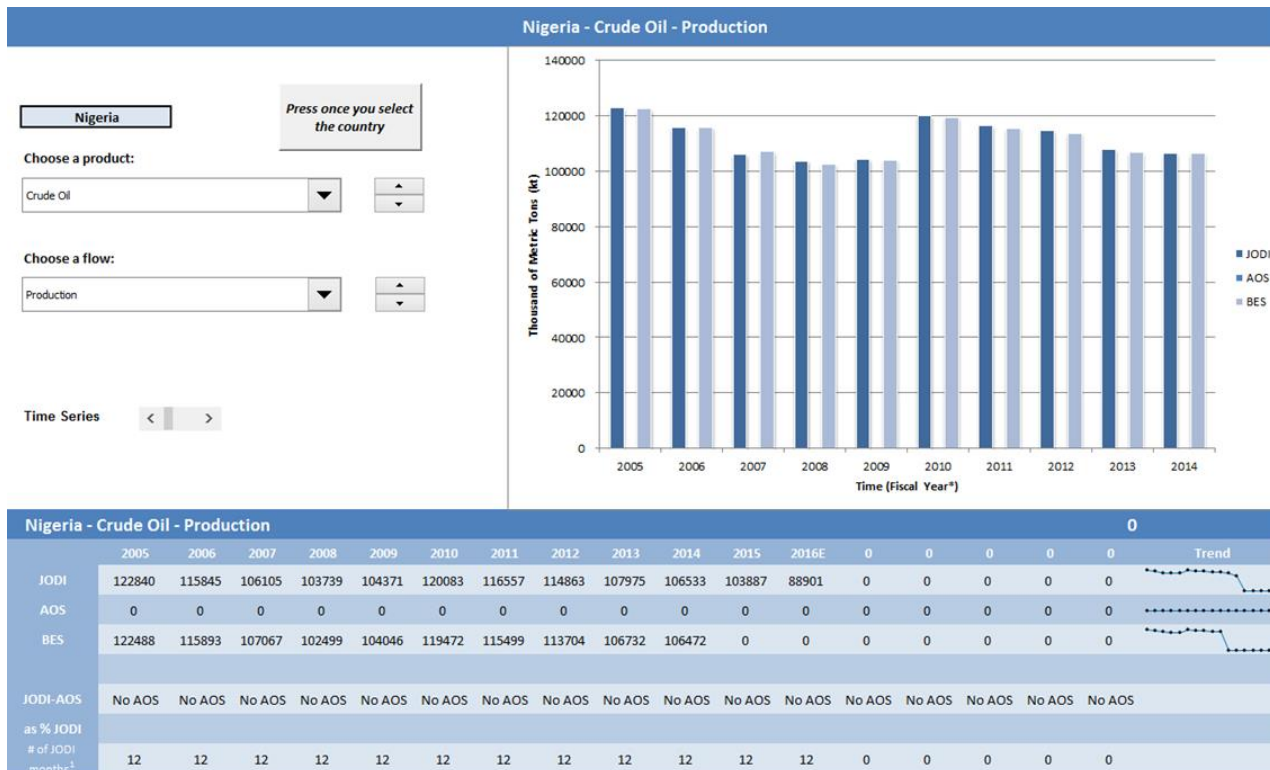
- JODI data are short-term data
- Availability and quality of early data
- Revisions to first released information
- Consistent methodology help make data useful
- Trends and levels

JODI data are official data



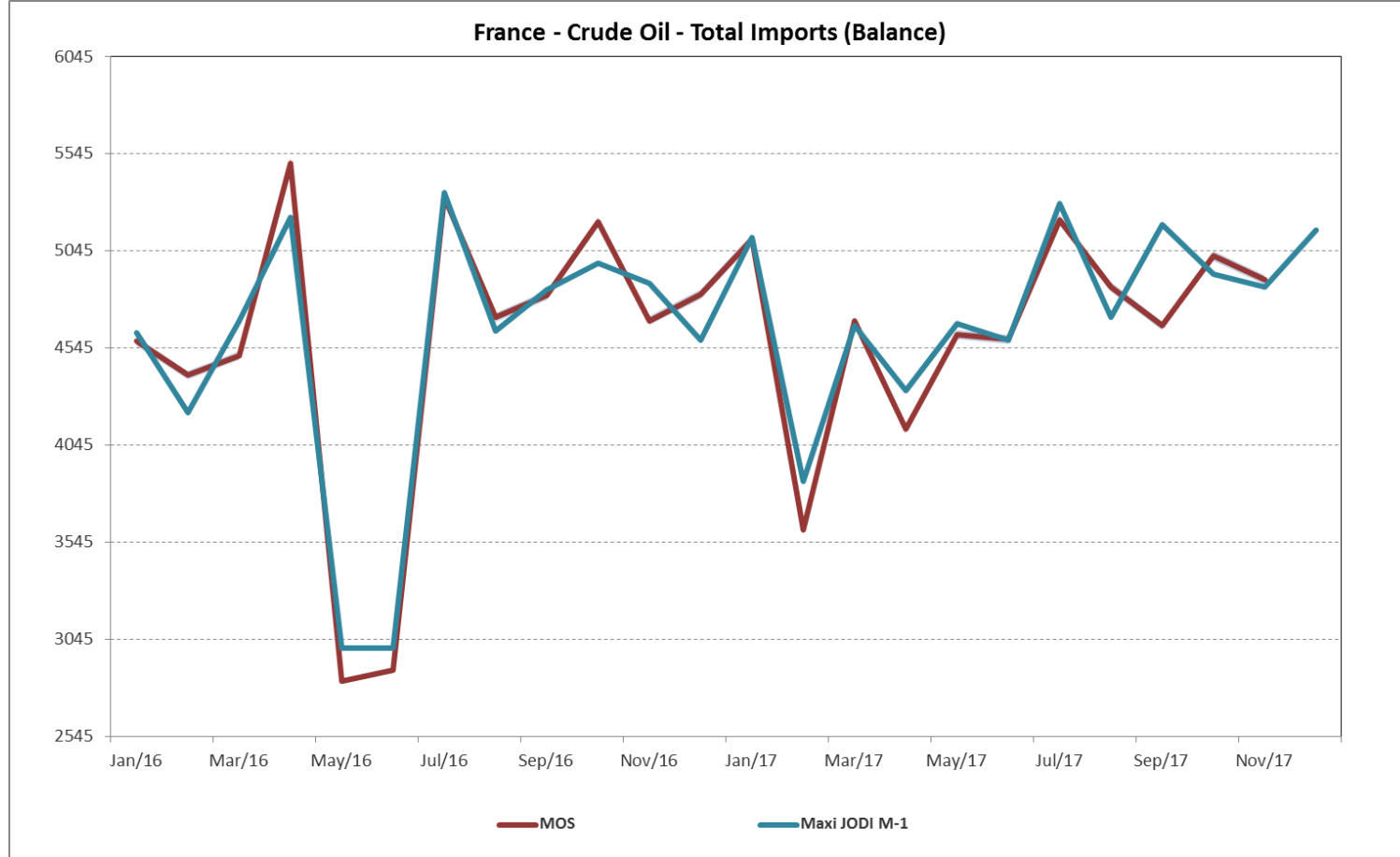
Assessing JODI Oil Data

- With annual data



1. Number of months per year in which JODI Maxi data are present

Checking vs other monthly data



Internal assessment of JODI Oil Data

JOINT OIL DATA INITIATIVE

APERC/EUROSTAT/OECD-IEA/OLADE/OPEC/UN

Country France

	Crude Oil	NGL	Other	Total		Petroleum Products									
						LPG	Naphtha	Gasoline	Total Kerosene	Of which: Jet Kerosene	Gas/Diesel Oil	Fuel Oil	Other Products	Total Products	
						M-1	M-1	M-1	M-1	M-1	M-1	M-1	M-1	M-1	M-1
+ Production					+ Refinery Output										
+ From Other sources					+ Receipts										
+ Imports					+ Imports										
- Exports					- Exports										
+ Products Transferred /Backflows					- Products Transferred										
- Direct Use					+ Interproduct Transfers										
- Stock Change					- Stock Change										
- Statistical Difference					- Statistical Difference										
= Refinery Intake					= Demand										
Closing stocks					Closing stocks										

Assessment - Based on comparisons with MOS data:

	Good
	Fair
	Less reliable

Assessment - Based on period: Dec-16 to Nov-17 for Dec-17

- Based on three indicators:

Sustainability of submission

Timeliness

Completeness

For the period July - December 2016:

Albania	☹️	😐	😊	Greece	☹️	😐	😊	Nigeria	☹️	😐	😊
Algeria	☹️	😐	😊	Grenada	n.a.	n.a.	n.a.	Norway	☹️	😐	😊
Angola	☹️	😐	😊	Guatemala	☹️	😐	😊	Oman	☹️	😐	😊
Argentina	☹️	😐	😊	Guyana	☹️	😐	😊	Panama	☹️	😐	😊
Armenia	☹️	😐	😊	Haiti	n.a.	n.a.	n.a.	Papua New Guinea	☹️	😐	😊
Australia	☹️	😐	😊	Honduras	n.a.	n.a.	n.a.	Paraguay	☹️	😐	😊

☹️☹️☹️ Sustainability (of submission) 😐😐😐 Timeliness 😊😊😊 Completeness

😊 Good 😐 Fair ☹️ Poor

Ongoing Evaluation (internally)



From January 2017 to June 2017
 Update to the marks for the JODI Oil data providers for the IEFS
 IEA/OECD countries (period From January 2017 to June 2017 data submissions) M-1

produce single page xlsx & pdf reports

better
worse

Country	Sustainab ility	Timeliness	Complete ness (%)	Sustainab ility	Timeliness	Complete ness (%)
Australia	6	6	100%	⊖	⊕	⊕
Austria	6	6	100%	⊖	⊕	⊕
Belgium	6	6	100%	⊖	⊕	⊕
Canada	6	6	100%	⊖	⊕	⊕
Chile	6	6	100%	⊖	⊕	⊕
Czech Republic	6	6	100%	⊖	⊕	⊕
Denmark	4	4	100%	⊖	⊕	⊕
Estonia	6	6	100%	⊖	⊕	⊕
Finland	6	6	100%	⊖	⊕	⊕
France	6	6	100%	⊖	⊕	⊕
Germany	6	6	100%	⊖	⊕	⊕
Greece	6	6	100%	⊖	⊕	⊕
Hungary	6	6	100%	⊖	⊕	⊕
Iceland	6	6	100%	⊖	⊕	⊕
Ireland	6	2	100%	⊖	⊕	⊕
Israel	0	0	0%	⊖	⊕	⊕
Italy	6	6	100%	⊖	⊕	⊕
Japan	6	6	98%	⊖	⊕	⊕
Korea	6	6	100%	⊖	⊕	⊕
Luxembourg	6	5	99%	⊖	⊕	⊕
Mexico	6	6	100%	⊖	⊕	⊕
Netherlands	6	6	100%	⊖	⊕	⊕
New Zealand	6	3	98%	⊖	⊕	⊕
Norway	6	6	100%	⊖	⊕	⊕
Poland	6	6	99%	⊖	⊕	⊕
Portugal	6	6	100%	⊖	⊕	⊕
Slovak Republic	6	5	100%	⊖	⊕	⊕
Slovenia	6	6	99%	⊖	⊕	⊕
Spain	6	6	100%	⊖	⊕	⊕
Sweden	6	6	100%	⊖	⊕	⊕
Switzerland	6	6	100%	⊖	⊕	⊕
Turkey	6	6	98%	⊖	⊕	⊕
United Kingdom	6	6	100%	⊖	⊕	⊕
United States	6	6	100%	⊖	⊕	⊕

From July 2016 to December 2016
 Update to the marks for the JODI Oil data providers for the IEFS
 IEA/OECD countries (period From July 2016 to December 2016 data submissions) M-1

Country	Sustainab ility	Timeliness	Complete ness (%)	Sustainab ility	Timeliness	Complete ness (%)
Australia	6	6	100%	⊖	⊕	⊕
Austria	6	6	100%	⊖	⊕	⊕
Belgium	6	6	100%	⊖	⊕	⊕
Canada	6	6	100%	⊖	⊕	⊕
Chile	6	5	94%	⊖	⊕	⊕
Czech Republic	6	6	100%	⊖	⊕	⊕
Denmark	6	6	100%	⊖	⊕	⊕
Estonia	6	6	100%	⊖	⊕	⊕
Finland	6	6	99%	⊖	⊕	⊕
France	6	6	100%	⊖	⊕	⊕
Germany	6	6	100%	⊖	⊕	⊕
Greece	6	6	100%	⊖	⊕	⊕
Hungary	6	6	100%	⊖	⊕	⊕
Iceland	6	6	100%	⊖	⊕	⊕
Ireland	6	1	100%	⊖	⊕	⊕
Israel	0	0	0%	⊖	⊕	⊕
Italy	6	6	100%	⊖	⊕	⊕
Japan	6	6	98%	⊖	⊕	⊕
Korea	6	6	100%	⊖	⊕	⊕
Luxembourg	6	6	100%	⊖	⊕	⊕
Mexico	6	6	100%	⊖	⊕	⊕
Netherlands	6	6	99%	⊖	⊕	⊕
New Zealand	2	1	100%	⊖	⊕	⊕
Norway	6	6	98%	⊖	⊕	⊕
Poland	6	6	100%	⊖	⊕	⊕
Portugal	6	6	100%	⊖	⊕	⊕
Slovak Republic	6	6	100%	⊖	⊕	⊕
Slovenia	6	6	99%	⊖	⊕	⊕
Spain	6	6	100%	⊖	⊕	⊕
Sweden	6	6	100%	⊖	⊕	⊕
Switzerland	6	6	100%	⊖	⊕	⊕
Turkey	6	6	99%	⊖	⊕	⊕
United Kingdom	6	6	100%	⊖	⊕	⊕
United States	6	6	100%	⊖	⊕	⊕

