



UPU | UNIVERSAL
POSTAL
UNION

UPU postal big data and trade

UN GWG on Big Data

Use cases for data, services and applications: trade data lake

10 November 2017



Outline

Digital trade context and international postal exchanges

International e-commerce growth through postal networks

UPU postal tracking data capture and the adoption of digital customs declaration systems

4V postal **data**, international postal flows and trade

Data on international postal shipments volumes and shifting digital trade patterns

Data on value and categories of goods traded online and trade in low-value shipments

UPU big data platform **technology**

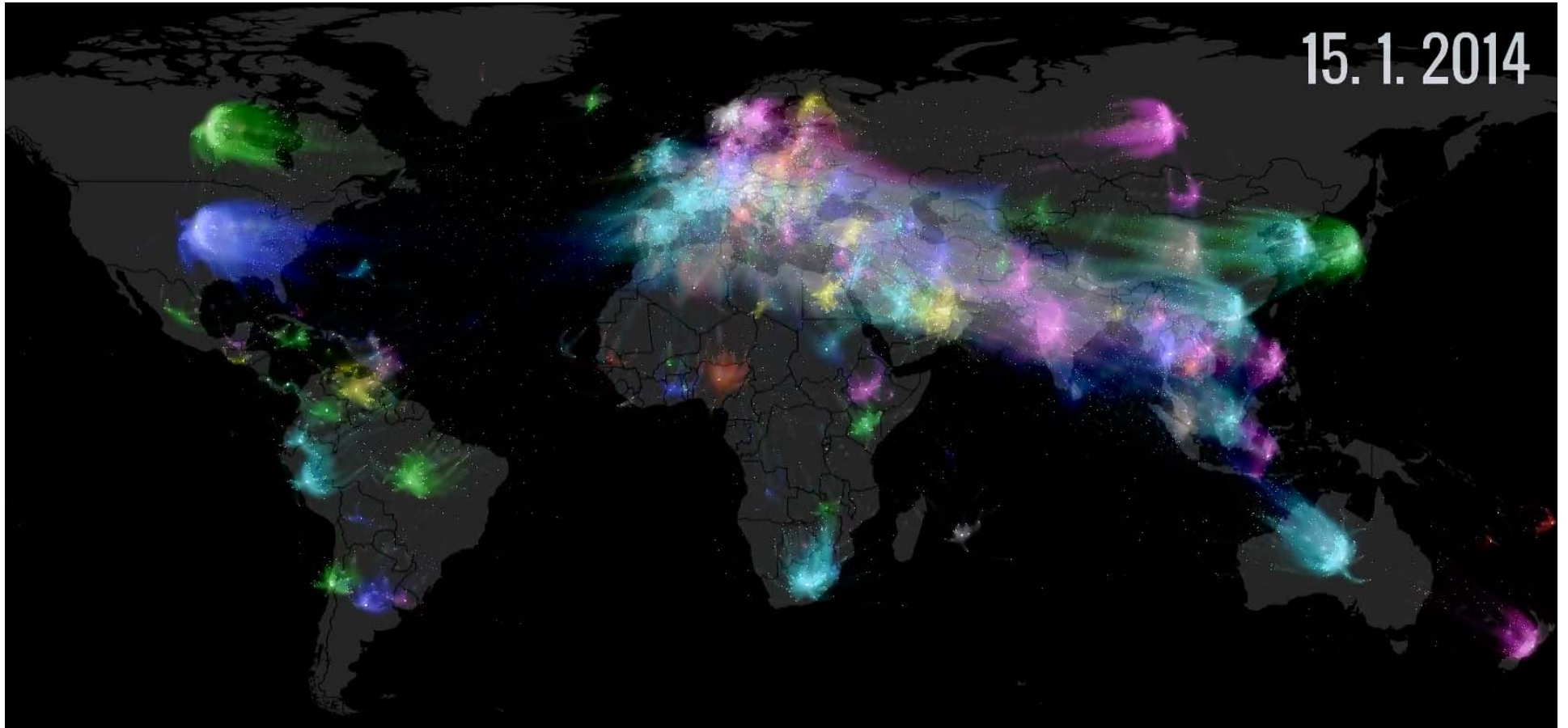
UPU big data zoo

Connecting to trade data lake

UPU big data **governance** and **collaborations**

Collaborations and cooperation agreement frameworks

Experience with UN Global Pulse





Quantitative aspects

Volume *Scale of data*

Over 30 billion historical international postal tracking records by 2020, exponentially growing since 2010 – 3 billion in 2016 alone

Velocity *Analysis of streaming data*

Hourly updates of millions of potential predicted delivery times and events along the international postal supply chain

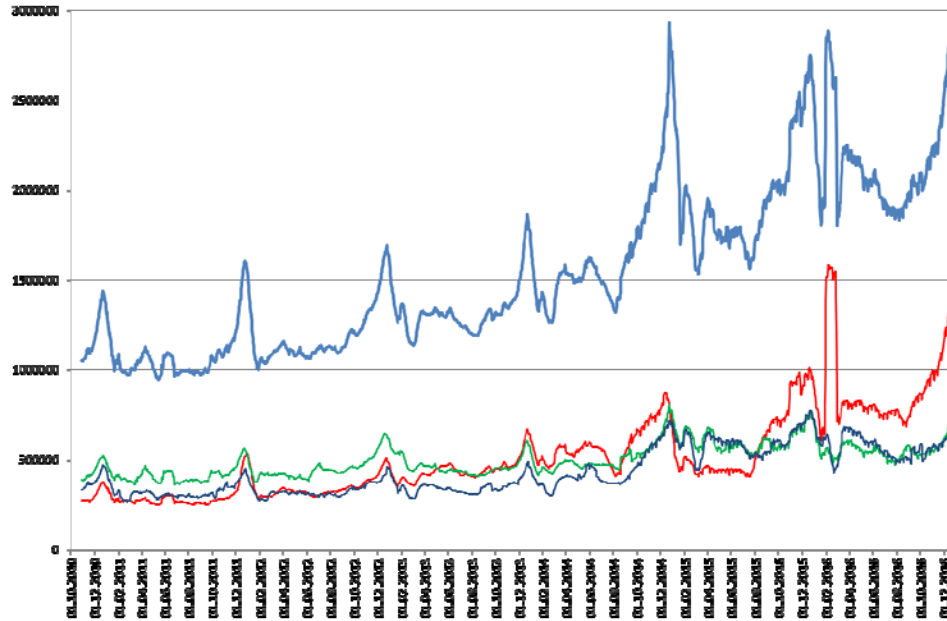
Qualitative aspects

Variety *Different forms of data sources*

Geo-located data (GIS standards, geocoded information), EDI postal tracking data (postal data standards), customs data, and aviation/transportation data

Veracity *Uncertainty of data*

Data reconciliation between all tracking data and sample-based QoS measurement system data (e.g. GMS), geographic coverage of different international postal data exchange systems



TONNAGE	total	letter-post	parcel-post	EMS
	2013	457'579'770	163'106'100	168'253'513
2014	566'648'650	213'610'159	188'587'382	164'451'109
2015	657'681'986	219'095'693	220'296'032	218'338'735
2016	742'928'846	332'658'171	203'304'623	209'583'027
2016/15	13.0%	51.8%	-7.7%	-4.0%
2015/14	16.1%	2.6%	16.8%	32.8%
2014/13	23.8%	31.0%	12.1%	30.3%

VOLUME	total	letter-post	parcel-post	EMS
	2013	396'747'127	318'548'512	38'290'757
2014	460'525'763	373'012'947	35'961'086	51'551'730
2015	569'431'727	461'646'446	40'404'849	67'380'432
2016	689'771'865	580'826'534	40'920'615	68'435'558
2016/15	21.1%	25.8%	1.3%	1.6%
2015/14	23.6%	23.8%	12.4%	30.7%
2014/13	16.1%	17.1%	-6.1%	29.2%



2011 (from/to)	Developed countries	Africa	Asia and Oceania	Latin America and Caribbean	Transition economies	World
Developed countries	46.3	2.4	12.1	7	2.8	70.6
Africa	0.7	0.2	0.1	0	0	1
Asia and Oceania	21.6	0.3	2.7	0.5	0.4	25.5
Latin America and Caribbean	1.7	0	0.1	0.3	0	2.1
Transition economies	0.5	0	0	0	0.3	0.8
World	70.8	2.9	15	7.8	3.5	100
2016 (from/to)	Developed countries	Africa	Asia and Oceania	Latin America and Caribbean	Transition economies	World
Developed countries	26.3	0.9	20.8	2.7	2.4	53.1
Africa	0.7	0.2	0.2	0	0	1.1
Asia and Oceania	33.2	0.4	4.2	1.4	4.0	43.2
Latin America and Caribbean	1.0	0	0.1	0.2	0	1.3
Transition economies	0.7	0	0.1	0	0.5	1.3
World	61.9	1.5	25.4	4.3	6.9	100

International deliveries (tonnage) of small packets, parcels and packages, 2011 and 2016, distribution of regional flows as a share of global flows, percent

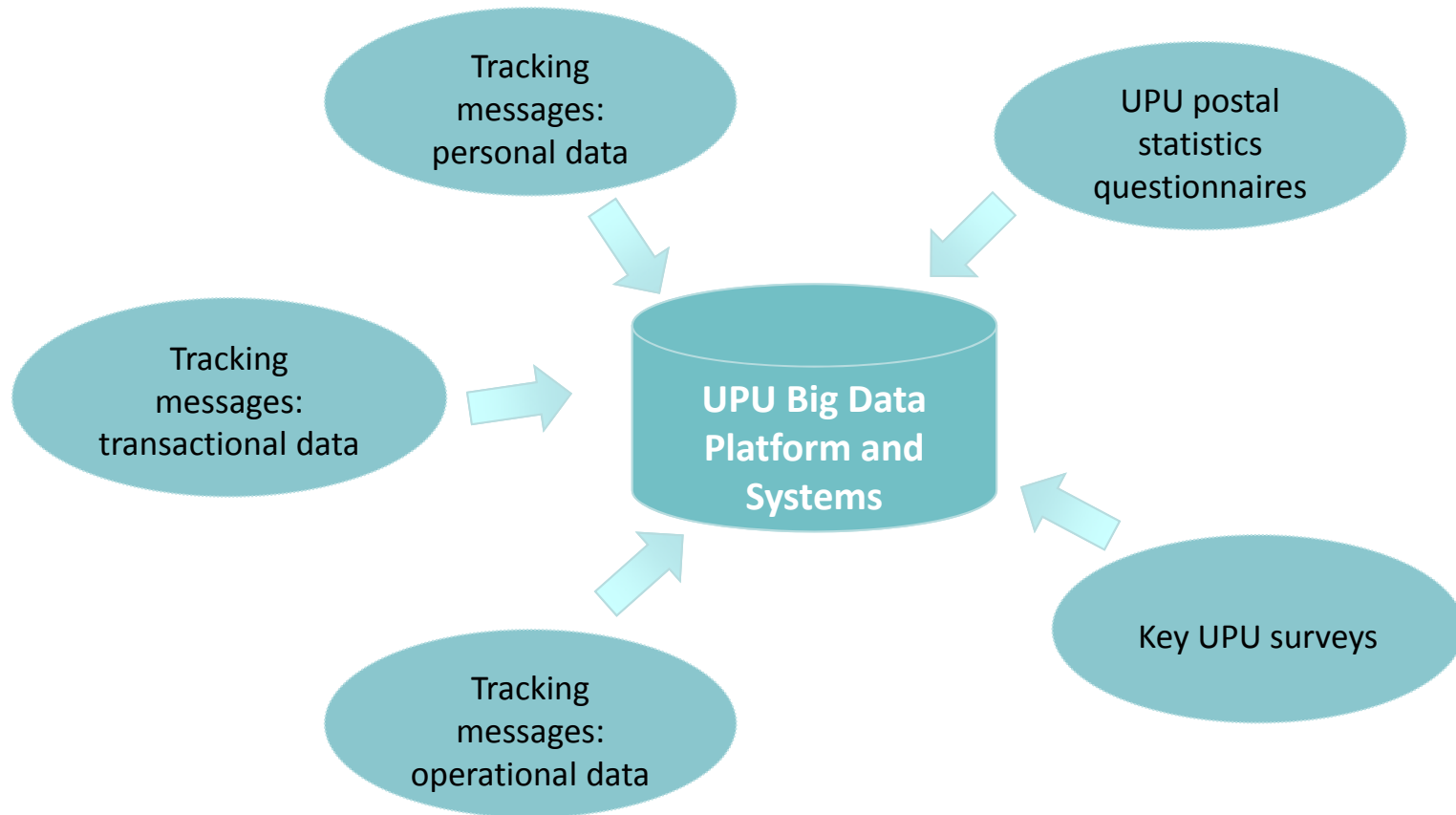
Source: Universal Postal Union.

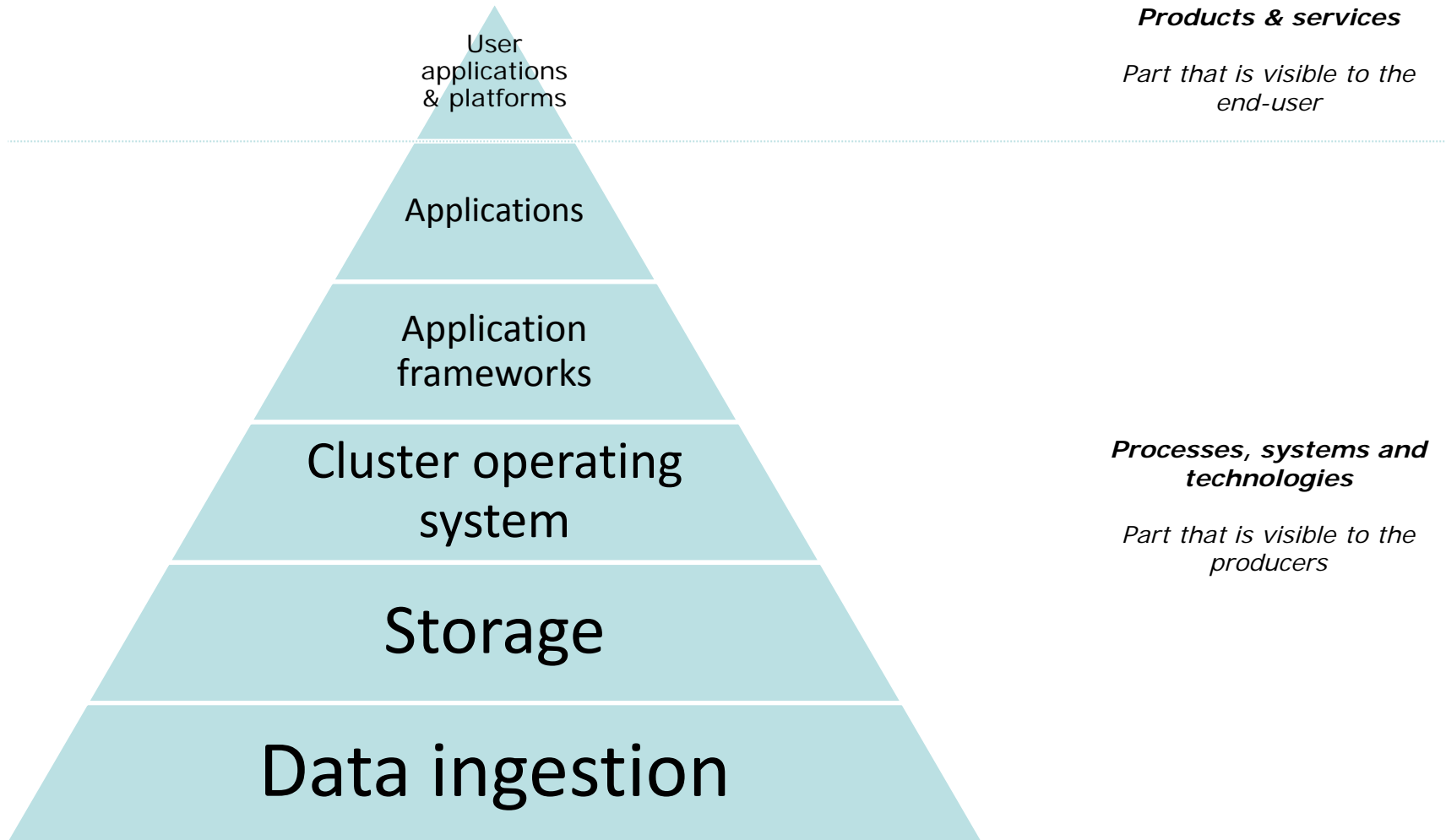


From postal volumes to international trade

**Currently: 379,301,707 electronic customs
declarations recorded in our UPU big data platform**

**285,952,139 with value less 50 (in any currency),
i.e. 75% of recorded declared postal shipments**







prod-bigd-edg03.ptc-prod.net:9995/#/notebook/2CT31KQN9

Zeppelin Notebook

/UPU/HBASE

```
import org.apache.spark.sql.functions.explode
val nm = mailitmatts.select(mailitmatts.col("id"), explode(mailitmatts.col("value")))
nm.createOrReplaceTempView("nm")

import org.apache.spark.sql.functions.explode
nm: org.apache.spark.sql.DataFrame = [id: string, key: string ... 1 more field]

Took 7 sec. Last updated by user1 at September 14 2017, 4:17:08 PM.
```

```
%spark2
nm.printSchema

root
 |-- id: string (nullable = true)
 |-- key: string (nullable = false)
 |-- value: string (nullable = true)

Took 1 sec. Last updated by faseld at September 14 2017, 2:55:06 PM. (outdated)
```

```
Filtering the exploded DataFrame for a specific key

Took 0 sec. Last updated by admin at September 19 2017, 4:22:23 PM.
```

```
%spark2
val df2 = sqlContext.sql("select cast(value as double) from nm where key = 'cp/d.val/amt'")
df2.count

df2: org.apache.spark.sql.DataFrame = [value: double]
res1: Long = 379301707

Took 33 min 0 sec. Last updated by user1 at September 14 2017, 4:50:19 PM.
```

Windows taskbar: prod-bigd-edg03.ptc... FRA 10:57 AM



UPU | UNIVERSAL
POSTAL
UNION

Thank you