D1.2 Progress report on Data Quality

in EU Member States

Grant Agreement number: 837354
Project acronym: SAVDON
Work Package number: WP1

Organisation: **World Marrow Donor Association (WMDA)**
LEAR: Esther Pustjens
Project coordinator: Lydia Foeken
Tel: 0031 88 505 7900
E-mail: lydia.foeken@wmda.info
Organisation website address: www.wmda.info

Table of Contents

Abbreviations

CBB = Cord Blood Bank

CBU = Cord Blood Unit

DD = Data Dictionary

DQ = Data Quality

HLA = Human Leucocyte Antigen

JSON = JavaScript Object Notation

QA = Quality Assurance

RFC = Request For Change

TNC = Total Nucleated Cell

WMDA = World Marrow Donor Association

XML = Extensible Mark-up Language

XSD = XML Schema Definition

1. Introduction

This progress report on data quality in EU Member States details the process and achievements each EU member state has made towards meeting the defined data set requirements in 2019. This report provides progress information to allow comparative analysis of progress by each EU member state and will enable them to focus their efforts on areas of identified weaknesses. In addition, this report includes tailored advice and guidance to each EU member to support them to remedy those areas of weaknesses.

This Deliverable D1.2 *Progress report on data quality in EU Member States* is part of the 2019 work programme of the World Marrow Donor Association (WMDA) for the EU Third Health Programme (2014-2020).

1.1 Background

In April 2018, WMDA successfully upgraded the data upload system for WMDA global donor and CBU database, which is used in the WMDA Search & Match Service. The new system uses the modern format XML instead of old format DOT20 in the upload file. After the system upgrade, majority of organisations (organisation means "donor registries or CBBs" in this document) gradually moved to use XML to upload. This transition period took almost one year. During this period, WMDA noticed that organisation encountered different issues, such as merging data, preparing valid XML files, correcting or cleaning up data based on WMDA business rules of data validation and HLA validation, etc. Any data processing can cause data quality issues. WMDA spent a lot of effort into strict Quality Assurance (QA) checks and controls during the data transition period until the organisations' data is good enough compared to the old DOT20 upload for production. Unfortunately, after this period, data quality issues as mentioned above arise again in consecutive uploads. In addition to this, data of some fields with less restrict validation raise issues as well, like Total number of nucleated cells post processing (TNC_FRZN). Moreover, the data upload system together with Search & Match system has weaknesses in data processing. The Data Quality (DQ) requirements and management become necessary after the transition period. Considering all the data process issue mentioned above, at end of 2018, WMDA planned and proposed the Data Quality project.

1.2 Requirements

The new XML data format significantly increases the number of data elements (fields) and adds the possibility to expand the definition to add more data elements in the future. More data may lead to more issues during processing and the data flow from the organisation side to WMDA Search & Match system. In order to guarantee the correctness and quality of the data as well as decrease the uncertainties for search coordinators when using Search & Match, WMDA defined a requirement plan to improve DQ based on 3 main data characteristics: Completeness, Timeliness and Reliability (Consistency and Accuracy). The DQ plan contains 2 scopes at the time when discussed and planned.

- Scope 1: Improve, optimize and enhance current data upload solution
- Scope 2: Investigate possible new solutions or replacement for certain parts in the data process flow for Search & Match

The details of the requirements are outlined on WMDA Share in the secure Working area of the Data Quality group. In 2019, WMDA mainly investigated and handled the requirements of Scope 1. *Table 1* shows an overview of the requirements that WMDA tackled.

Table 1. Overview of WMDA 2019 DQ requirements

Requirements	Working/Implementation Period	Data Characteristic(s)
Monthly data upload	January 2019 – July 2019	Timeliness
Density Improvement Plan	January 2019 – December 2019	Completeness
Deprecated Code	November 2018 – March 2019	Reliability
Data upload Dashboard and Statistic Dashboard	January 2019 – December 2019	Data monitor /track Purpose
Record Drop Control 10% Rule	June 2019	Reliability Completeness
Differential Upload	May 2019 – September 2019	Timeliness
Density report per registry	April 2019 – December 2019	Data monitor /track Purpose
XSD 2.2	August 2019 – December 2019	Completeness Reliability

Since most of the small organisations do not have a data quality management system within their organisation, the DQ requirements plan helps these small organisations to improve their data internally.

In Chapter 2. Progress and Achievements, details how WMDA handles these requirements are provided.

2. Progress and Achievement of Data Quality

2.1 Monthly data upload

After most of the donor registries and CBBs moved to the implementation of XML in 2018, WMDA introduced the requirement that uploads should be done at least on a monthly basis. In 2019, this requirement was added to the WMDA Standards 2020, which makes it mandatory for WMDA Qualified and Accredited organisations. In the first half year of 2019, based on the data upload track per registry in the data upload dashboard, WMDA contacted all registries and CBBs that did not update monthly. Thanks to their hard work, the upload frequencies were gradually increasing in the first half of 2019 and from July 2019 the monthly file upload number remains at about 600 (*Figure 1*).

Figure 1. Monthly data upload track chart for CBU, Donor and Total
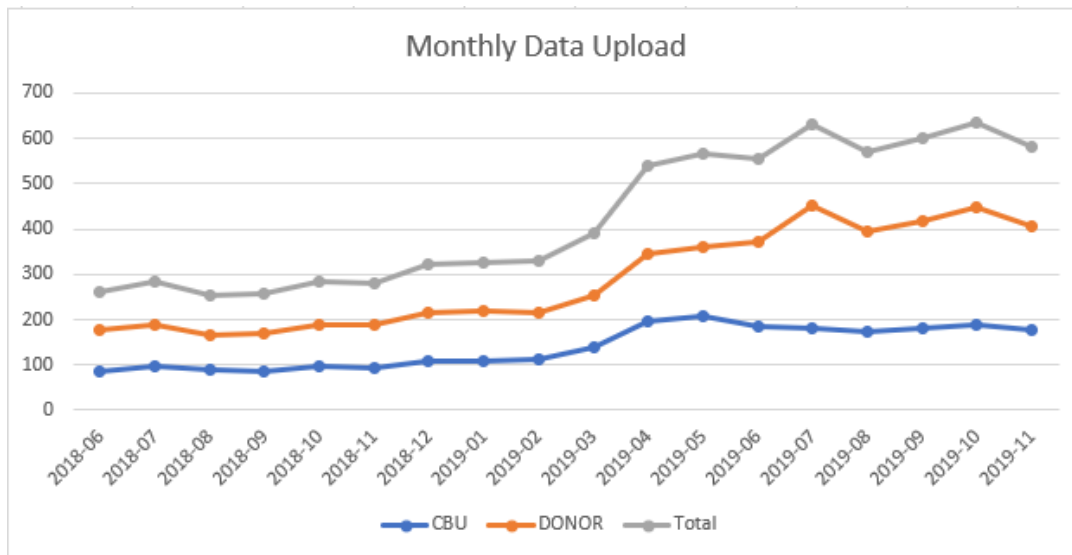
Until December 2019, the distribution of the XML upload frequency is as shown in *Figure 2*.

Figure 2. The Upload Frequency Distribution by Organisation and Records (DONOR and CBU)
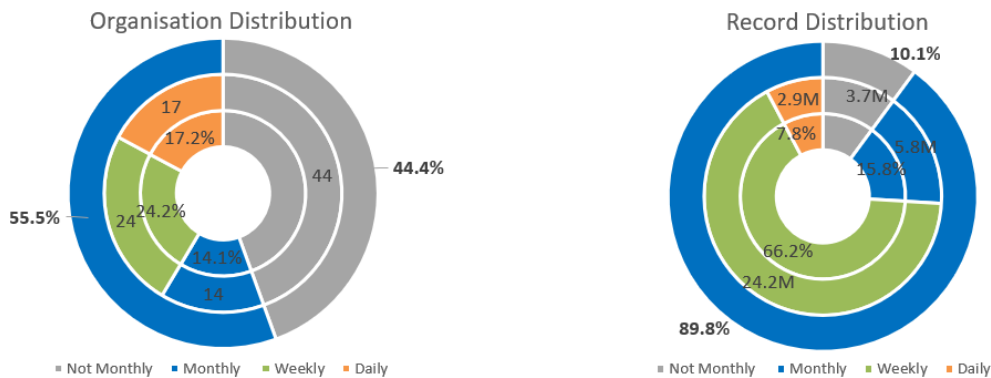
WMDA noticed that 44.4% of the organisations did not achieve the WMDA requirements. Most of those are CBBs. The reason for this is because Cord Blood Unit (CBU) collection and release gradually decreased, and the updates are less frequent. On the positive side: in total almost 90% of the records are uploaded monthly (15.8%), the majority are uploaded weekly (66.2%) and some even daily (7.8%).

For all WMDA EU members, the upload frequency is as shown in *Figure 3*. About 74.2% organisations upload monthly or weekly or daily, these contribute 95.7% of the records.

Figure 3. The EU Upload Frequency Distribution by Organisation and Records (DONOR and CBU)

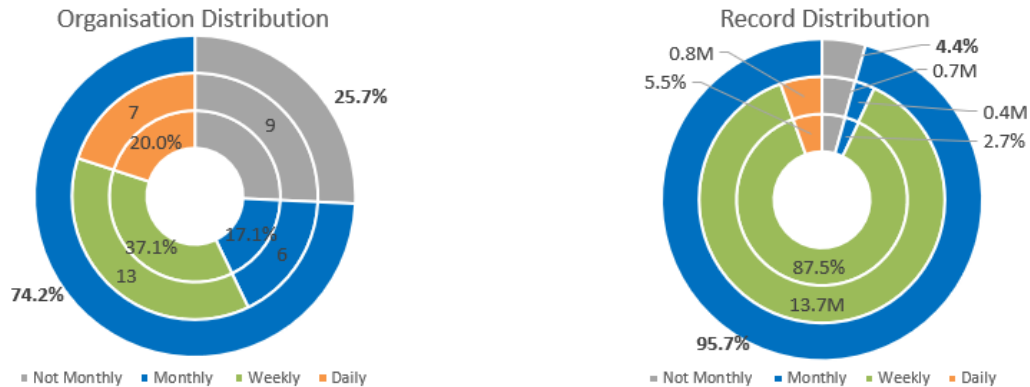The requirement to upload the dataset on a monthly basis was a starting point to improve the timeliness of the data. For donor registries who have millions of donors, donor data gets updated daily. In order to close the gap between WMDA global donor database compared to registries' local database, WMDA investigated and planned the Differential Upload (Details in 2.6) subproject.

After the feature be released by WMDA and used by registries, WMDA foresees that the weekly or daily upload will increase. The differential upload allows data providers to upload only the changed or new records instead of uploading the full dataset with all records again. This will allow for faster and smaller uploads, lowered system loads and faster quality checks.

2.2 Density Improvement Plan

At the end of 2018, WMDA introduced the Density Improvement Plan for 2019. For this, WMDA identified a first set of data elements with high clinical impact. These elements are in the first focus group for improvement. The density describes the percentage of records that actually has data available in that specific field, so this is a quantitative metric. In *Figure 4* the average density of end of 2018 and end of 2019 as writing now is provided.

Figure 4. Density improvement plan for 2019

Attention Time	XML Field	Data Type	XML Field Description	Percentage 2018	Percentage 2019	Status
2019-02	ETHN	DONOR	Ethnicity	30%	56%	Approved
2019-02	SEX	DONOR	Sex	95%	96%	Approved
2019-02	ABO	DONOR	ABO	38%	43%	Approved
2019-02	BIRTH_DATE	DONOR	Age	95%	96%	Approved
2019-06	CONTACT_DATE	DONOR	Last contact date	33%	36%	Approved
2019-10	CMV	DONOR	CMV status	9%	12%	Approved
2019-04	NMBR_TRANS	DONOR	No. of donations	6%	4%	Approved
2019-02	BANK_MANUF_ID	CBU	Accreditation Status	67%	76%	Approved
2019-02	COLL_DATE	CBU	Collection date	89%	57%	Approved
2019-02	SEX	CBU	Sex	78%	80%	Approved
2019-02	ABO	CBU	ABO	73%	73%	Approved
2019-04	TNC_FRZN	CBU	post processing TNC	92%	98%	Approved
2019-04	CD34PC_FRZN	CBU	Post processing CD34+	69%	73%	Approved
2019-02	ETHN	CBU	Ethnicity	37%	40%	Approved
2019-05	VOL_FRZN	CBU	post processing volume	82%	81%	Approved
2019-09	VIABILITY	CBU	Viability	36%	38%	Approved
2019-10	MNC_FRZN	CBU	post processing MNC	22%	21%	Approved
2019-11	ATT_SEG	CBU	Attached segments	33%	32%	Approved

In order to work on this plan, WMDA created 2 dashboards for organisations on WMDA Share:
1) The last 90 days global average density to track the average value and changes of the density for all the fields (*Figure 5*)
2) The "Fields of This Month" dashboard in the Data Upload Dashboard to track all the fields per organisation (*Figure 6*)
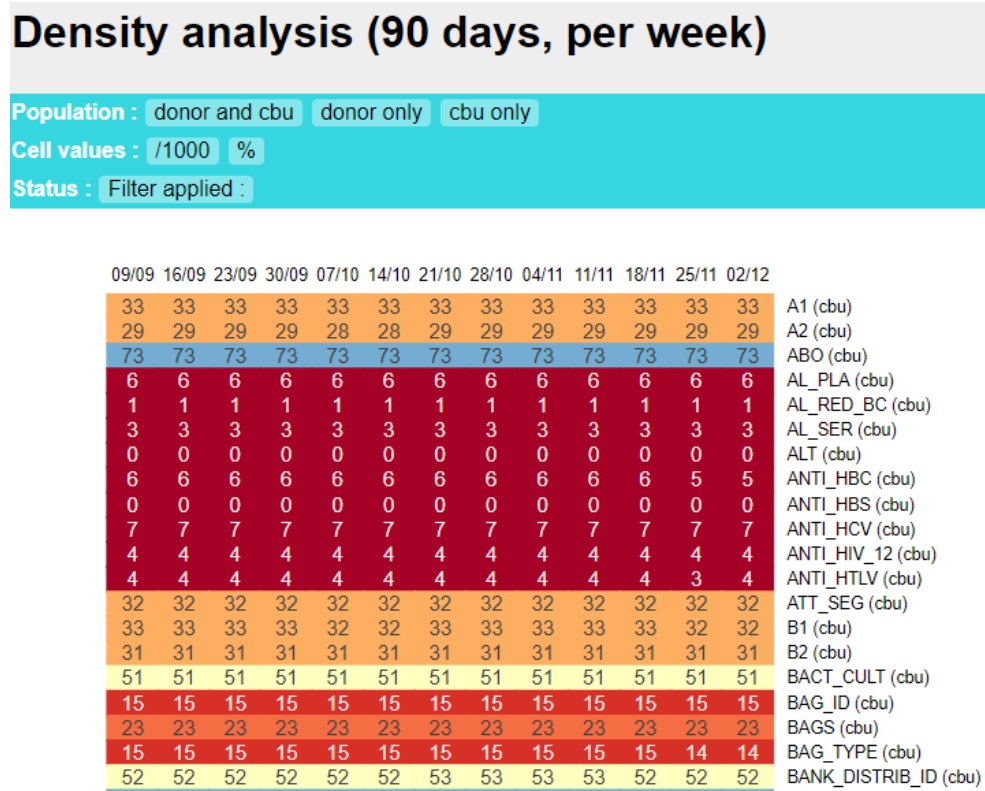
Figure 5. The 90 days global average density report

Density analysis (90 days, per week)

Population : donor and cbu | donor only | cbu only
Cell values : /1000 | %
Status : Filter applied :

	09/09	16/09	23/09	30/09	07/10	14/10	21/10	28/10	04/11	11/11	18/11	25/11	02/12	
	33	33	33	33	33	33	33	33	33	33	33	33	33	A1 (cbu)
	29	29	29	29	28	28	29	29	29	29	29	29	29	A2 (cbu)
	73	73	73	73	73	73	73	73	73	73	73	73	73	ABO (cbu)
	6	6	6	6	6	6	6	6	6	6	6	6	6	AL_PLA (cbu)
	1	1	1	1	1	1	1	1	1	1	1	1	1	AL_RED_BC (cbu)
	3	3	3	3	3	3	3	3	3	3	3	3	3	AL_SER (cbu)
	0	0	0	0	0	0	0	0	0	0	0	0	0	ALT (cbu)
	6	6	6	6	6	6	6	6	6	6	6	5	5	ANTI_HBC (cbu)
	0	0	0	0	0	0	0	0	0	0	0	0	0	ANTI_HBS (cbu)
	7	7	7	7	7	7	7	7	7	7	7	7	7	ANTI_HCV (cbu)
	4	4	4	4	4	4	4	4	4	4	4	4	4	ANTI_HIV_12 (cbu)
	4	4	4	4	4	4	4	4	4	4	4	3	4	ANTI_HTLV (cbu)
	32	32	32	32	32	32	32	32	32	32	32	32	32	ATT_SEG (cbu)
	33	33	33	33	32	32	33	33	33	33	33	32	32	B1 (cbu)
	31	31	31	31	31	31	31	31	31	31	31	31	31	B2 (cbu)
	51	51	51	51	51	51	51	51	51	51	51	51	51	BACT_CULT (cbu)
	15	15	15	15	15	15	15	15	15	15	15	15	15	BAG_ID (cbu)
	23	23	23	23	23	23	23	23	23	23	23	23	23	BAGS (cbu)
	15	15	15	15	15	15	15	15	15	15	15	14	14	BAG_TYPE (cbu)
	52	52	52	52	52	52	53	53	53	53	52	52	52	BANK_DISTRIB_ID (cbu)

Figure 6. Example of the density report per registry

Name	Listing Org	ION	Field	Total Records	Filled Records	Percentage	Created_at
REDOME - Registro Nacional de Doadores Voluntarios de Medula Ossea	8766	8766	TNC_FRZN	16840	16136	95.8194774346793	2019-12-03 15:51:01
NHS Blood and Transplant - BBMR	2731	2731	TNC_FRZN	18337	18310	99.852756721383	2019-12-03 12:08:32
Austrian Bone Marrow Donor Registry	2614	2614	TNC_FRZN	83	82	98.7951807228916	2019-12-02 19:00:43
Finnish Stem Cell Registry	9738	9738	TNC_FRZN	3340	3332	99.7604790419162	2019-12-02 18:56:48
Tobias Registry	5285	5285	TNC_FRZN	4356	3508	80.5325987144169	2019-12-02 18:55:07
Slovenia Donor	4565	4565	TNC_FRZN	342	339	99.1228070175439	2019-12-02 18:40:40

Based on these dashboards, WMDA reached out to most of the organisations to check whether they can provide these fields or can increase the density of these fields to average level of WMDA global database. More information and progress related to these dashboards is provided in section 2.4.

Table 2 provides the density of the focused fields with EU vs Non-EU. Please note that not all fields are available with data for both donors or CBUs. A full density comparison report of EU vs Non-EU can be found in [Appendix 1.](#)

Table 2. Density report of EU members VS non-EU members in focused fields

Density (data available) in %				
	Non-EU	EU member states	Non-EU	EU member states
Data element	Cord Blood Units	Cord Blood Units	Donors	Donors
ABO	69.2	73.0	44.6	51.5
ATT_SEG	15.6	15.0		
BIRTH_DATE	81.3	100.0	91.3	100.0
BANK_MANUF_ID	60.7	75.0		
CD34PC_FRZN	68.2	80.6		
CMV	12.0	11.5	4.6	16.2
COLL_DATE	36.7	33.0		
CONTACT_DATE			7.1	24.0
ETHN	20.8	29.7	35.4	34.0
GRID			89.0	94.3
NMBR_TRANS			14.6	35.3
SEX	60.8	90.1	93.5	100.0
TNC_FRZN	98.1	94.3		
VIABILITY	16.7	19.7		
VOL_FRZN	88.4	90.6		

During the density improvement process, further investigation has been done on the value and the correctness of specific fields since this value is crucial in the matching algorithm used for WMDA's Search & Match. Therefore, more time was spent on the value-distribution report. This qualitative report shows a more detailed insight in the dataset. This helps organisations to further improve key indicators by creating additional validations during pre-processing. It also gives the possibility for WMDA to provide tailored improvement suggestions and best practices to the organisations. In order to give a better explanation how this works, an example of TNC_FRZN improvement process is provided.

Example: TNC_FRZN (Total Nucleated Cells in CBU post processing/before freezing)

TNC_FRZN is an important parameter in finding a matching CBU, next to HLA. Therefore, WMDA spent a lot of effort to help CBBs to improve not only the density, but also the value for this field. From the density improvement plan table in *Figure 6*, it is clear that TNC_FRZN density increased from 92% to 98%. Due to the importance of the actual value of TNC_FRZN, search coordinators want to make sure that the values provided are correct. To assess the correctness of these values, WMDA started to do a qualitative analysis of TNC_FRZN to identify possible anomalies. Based on the TNC_FRZN distribution report of each organisation, CBBs were notified about their performance and values, and methods and suggestions for improvement were discussed.

Figure 6. The Real-Time Mean TNC_FRZN distribution per EU CBBs

Realtime Mean TNC_FRZN/TNC by Organizations

Name	ION	Total records	Count	Mean of TNC_FRZN	Mean of TNC
Anthony Nolan London	6354	8378	8378	134.58	null
Marrow Donor Program Registry Belgium	4201	21989	21536	121.44	145.84
Hellenic Cord Blood Bank	4868	2884	2884	109.74	null
Registro de Donantes de Médula Ósea (REDMO)	7813	63930	63894	125.38	138.18
NHS Blood and Transplant - BBMR	2731	18353	18326	135.39	180.27
France Greffe de Moelle Registry - FGM	1804	36890	36890	142.62	null
Czech Stem Cells Registry Prague (CSCR)	4753	4195	4194	120.77	122.84
Matchis Foundation	8139	4706	4706	118.23	null
Poltransplant	5391	3295	3146	102.54	null
ZKRD - Zentrales Knochenmarkspender-Register Deutschland	6939	17840	16964	99.35	129.14
José Carreras Cord Blood Bank Düsseldorf-Universitätsklinikum Düsseldorf	4908	20398	20398	125.96	null
Italian Bone Marrow Donor Registry	7450	38108	38092	130.24	null
Blutspende SRK Schweiz AG	9341	4912	4912	117.7	null
EOM Hellenic Transplant Organization	4979	77	77	88.19	null
Cyprus Bone Marrow Donor Registry	9751	3410	3321	122.96	126.22
Croatian Bone Marrow Donor Registry	5712	3575	0	null	null
Bulgarian Bone Marrow Donor Registry	7197	284	284	91.11	null
Thessaloniki Public Cord Blood Bank	6121	1555	1555	88.87	null
Slovenia Donor	4565	342	339	86.09	70.0
Austrian Bone Marrow Donor Registry	2614	83	82	168.53	161.8
Slovenský register placentárnych krvotvorných buniek- Eurocord-Slovakia (SRPKB)	4398	1693	1693	187.87	null

Below is a list of the fields similar to the situation as TNC_FRZN that WMDA identified as candidates for improvement, both on density as qualitative aspects. These fields were included in the communication with organisations and the corresponding improvement plans (*Table 3*).

Table 3. List of fields that are qualitatively monitored for CBU

Field name	Explanation	Normal Value Range
TNC_FRZN	Total Nucleated Cells in CBU post processing/prior to cryopreservation,	50×10^7 to 300×10^7
CD34PC_FRZN	Total number of CD34+ cells (post processing, prior to cryopreservation),	50×10^7 to 300×10^7
VOL_FRZN	Total volume frozen (post processing, prior to cryopreservation) in ml.	About 25 or 50
VIABILITY	A calculated score based on specific test in %	Between 80-100

2.3 Deprecated Code

HLA coding used in the data upload is validated by HLA validation engine. Since the nature of some of these codes is volatile based on quarterly HLA nomenclature update, codes may get deprecated. After a certain time, normally a one-year peace period, deprecated codes cannot be accepted anymore due to incompatibilities in the matching process. For this situation automatic code improvement suggestions in the processing reports were introduced. This enables data uploaders to keep HLA data of donors or CBBs up to date and relatively correct. A User Guide in public "Search & Match Service Data submission information" page is provided on how to process deprecated code. Below in *Table 4* is an example of report generated during data upload validation. The new valid code is suggested if it is available.

Table 4. The example report of the Deprecated code recommendation

```
W | 0999 | 100000100 | HLA: #Warning at locus A*: deprecated HLA code A*23:AMHJE is now invalid and passed its grace period. Locus blanked. New: 23:AMGRE
W | 0999 | 100000101 | HLA: #Warning at locus B*: deprecated HLA code B*13:DAZT is now invalid and passed its grace period. Locus blanked. New: 13:CXGS
W | 0999 | 100000102 | HLA: #Warning at locus A*: deprecated HLA code A*11:GNMX is now invalid and passed its grace period. Locus blanked. New: 11:NVYZ
W | 0999 | 100000103 | HLA: #Warning at locus A*: deprecated HLA code A*23:DFTW is now invalid and passed its grace period. Locus blanked. New: 23:AFVYZ
```

2.4 Data upload Dashboard and Statistic Analysis

In order to handle the Monthly data upload and Density Improvement Plan, real time dashboards that provide concise and comprehensible information are essential. Several dashboards were created to provide up to date information on data quality related parameters. Where possible, WMDA provides the data in an open format - JavaScript Object Notation (JSON) - for electronic processing by involved parties.

Name	Access	Description	Link
Data Upload dashboard	WMDA members	Upload statistics, full density report per each organisation	See *Figure 7* below
Current data availability	WMDA members	Number of records, typing, and recency of the dataset. Includes accreditation status.	https://statistics.wmda.info
Statistical Analysis for value Distribution	Public	Summarised and detail qualitative analysis of specified fields for each organisation	See *Figure 8* below
90 days average density	public	Trend analysis on global availability of data elements	https://statistics.wmda.info/density/
(Monthly) Trend analysis	public	Provide the last 30 days' number of records for each registry and the changes of percentage	https://statistics.wmda.info/spark/
Internal upload log dashboard	WMDA Internal	Provide the information of the events of each file upload, like warning, failed, xsd rejection, etc.	See *Figure 9* below

Figure 7. The data upload dashboard

Data Upload Dashboard

Created by Zhihong He, last modified just a moment ago

GRID in WSMS | **Data Upload Frequency** | Fields of This Month | Upload Overview | TNC Statistics

ⓘ **Dash usage guide**
1. Filter usage:
   - You can click on the "=" (equal/include) to make it "!=" (not equal/exclude) to select the value you want.
   - In pivot table, you can select value "0" to filter to get the orgs that have no upload in the last 1 or 2 or 3 months as you prefer.
   - You can add more filters or change the pivot table on the fly if you are an administrator to adjust the table as you want. But pls do not save.
2. Data explanation:
   - If you can not find an ION, then their data still in dot20 and no XML is uploaded yet. We are working to help these registries or banks.
   - Considering data size, only **last 18 months'** upload are calculated.
3. Detail source table can be found here. https://share.wmda.info/x/4g6OF

Data is updated every hour. **Last data update**: 2019-12-12 16:41:00+01

ION_NAME = [4201 Marrow Donor Program Registry Belgium] | Type = Click and start typing... | Global Filter = Start typing... | Count 2019-09 = from Type value to Type value | Count 2019-10 = from Type value to Type value

ION_NAME, Type / Month		Count																			
		2018-06	2018-07	2018-08	2018-09	2018-10	2018-11	2018-12	2019-01	2019-02	2019-03	2019-04	2019-05	2019-06	2019-07	2019-08	2019-09	2019-10	2019-11	2019-12	Total
⌄ 1726 Welsh Bone Marrow Donor Registry	D	1	1	1	1	1	1	1	1	2	4	5	4	4	5	4	4	5	4	2	51
	Total	1	1	1	1	1	1	1	1	2	4	5	4	4	5	4	4	5	4	2	51
⌄ 2015 Danish Stem Cell Donors - West	D	2	1	1	1	1	3	4	4	4	13	22	23	19	23	19	21	20	19	8	208
	Total	2	1	1	1	1	3	4	4	4	13	22	23	19	23	19	21	20	19	8	208
⌄ 2073 Lithuanian National Bone Marrow Donor Registry	D		1				3	4	2	4	7	22	23	19	23	20	21	23	20	8	200
	Total		1				3	4	2	4	7	22	23	19	23	20	21	23	20	8	200

Figure 8. Data upload statistical analysis dashboard

Data upload Analysis

Created by Zhihong He, last modified about 9 hours ago

TNC Statistics | TNC_FRNZ per org | CCR5 | HLA

All TNC_FRZN and TNC value is with factor x10^7.
Data is updated every hour. **Last data update**: 2019-12-16 09:42:00+01

Issue list with TNC_FRZN>500*10^7

❯ Click here to expand...

Data is till 2018-12-31. * (see note below)

TNC Overview

Total Records	Count	Mean TNC_FRZN	Mean TNC
709471	635164	114.77	129.15

Data is updated every hour. **Last data update**: 2019-12-16 09:42:00+01

TNC Overview

Total Records	Count	Mean TNC_FRZN	Mean TNC
789689	776188	115.58	141.41

TNC _FRZN categories

[-1,0] ■ [1,49] ■ [50,89] ■ [90,124] ■ [125,149]
[150,199] ■ [200,249] ■ [250,299] ■ [300,3000] ■ null

TNC _FRZN categories

[-1,0] ■ [1,49] ■ [50,89] ■ [90,124] ■ [125,149]
[150,199] ■ [200,249] ■ [250,299] ■ [300,599]
[600,3000] ■ null

Figure 9. Internal upload log dashboard

PRODUCTION : Last Events Pipelog

(Last modified: December 12 2019 11:45:01. Europe/Berlin +01:00 CET) [Reverse order] csv

Filter: error | DIFF records processed error warning upload × clear filter

Upl.ID	Time	ION	Name	Description	Code	Type	File
9518	2019-12-10 18:20:18	XXXX	Bone Marrow Donor Registry	There is an XML structure error on line 1: Element 'NMBR_MARR': This element is not expected. Expected is one of (RSV_PAT, STATUS). There is an XML structure error on line 1: Element 'NMBR_TRANS': This element is not expected. Expected is one of (RSV_PAT, STATUS).	xsd_rejected	xsd	ION-XXXX-D.xml.pgp
9493	2019-12-09 18:30:18	XXXX	Bone Marrow Donor Registry	There is an XML structure error on line 1: Element 'NMBR_MARR': This element is not expected. Expected is one of (RSV_PAT, STATUS). There is an XML structure error on line 1: Element 'NMBR_TRANS': This element is not expected. Expected is one of (RSV_PAT, STATUS).	xsd_rejected	xsd	ION-xxx-D.xml.pgp
9490	2019-12-09 12:50:12	9999	Marrow Registry	File could not be processed due to a file decryption error. Please check log file for error details.	failed	file	ION-xxx-D.xml.pgp

2.5 Record Drop control 10% rule

The WMDA data upload system is the data-ingress processing unit, that receives the data, decrypts it, decodes it, stores it and performs the validations to ensure the correctness of the provided data. It has been designed to monitor abrupt changes in the quantity of records during the XML data upload. This feature was not yet available in the first released version of the data upload system. When WMDA noticed that there was an abrupt change of more than 10%, based on the trend report, this early warning mechanism has been activated.
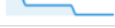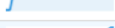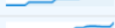
When a sudden or unexpected drop in the number of records is noticed, WMDA sends a notification out to organisations to check their data and holds the processing. WMDA will send congratulations to organisatio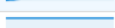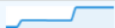ns that show a big increase in provided records within a month (*Figure 10*). WMDA provides this global trend information publicly at https://statistics.wmda.info/spark/ .

Figure 10. Trend report for last 1 month (2019)

Trend analysis GCD2 : WMDA Number of Records

source data (JSON) Last modified:December 03 2019 05:55:01.

Filter: search this table

ION	Pool	Type	Trend	Current ▲	Delta	Delta %
Total		Donor		35,809,009	112185	1
6939	6939	Donor		8,954,076	30480	1
3553	3553	Donor		8,852,871	-17821	0
8766	8766	Donor		4,663,389	26065	1
6939	7414	Donor		1,519,296	13473	1
4987	4987	Donor		1,001,941	3228	1
2197	2197	Donor		993,145		0
Total		CBU		789,054	5669	1
6354	6354	Donor		780,237	2879	1

2.6 Differential Upload

Currently, only full datasets from organisations are allowed to be uploaded in the WMDA data upload system. Considering only small amounts of donors / CBUs are actually changed, added or deleted daily or weekly, this is a waste for file preparation and process, both in capacity for validation, data transfer, and time. The differentials upload aims to set up a method to allow for only partial data-uploads that contain only the differentials between the old and the new dataset.

In order to clarify the demand of this feature and have an overview of the preference of organisations, WMDA carried out a Differential Upload Survey.

The survey results indicated that the majority welcomed this feature, and especially the larger organisations (with high volume datasets) were in favour to use this mechanism, since it allows them to improve the timeliness of their dataset, by pushing just the differences. Currently, the differential feature is deployed in data upload staging system for test purpose. WMDA plans to have it tested by 1-2 registries with more than one million records in the dataset. If this is successful, it will be moved to the production environment.

2.7 Density report per registry

During the DQ process, WMDA received requests from organisations that they prefer to check whether the data upload to WMDA after processing is consistent with the data in their own system. At the same time, WMDA would like each registry to assess the quality of the data periodically, without explicit notification by WMDA. This means it is necessary to distribute the density report directly to the organisations.

2.7.1 Example density

Each organisation will receive a tailored and organisation specific report that provides progress reporting on key indicators. These reports are currently being reviewed; therefore a sample report is provided in *Figure 11*.
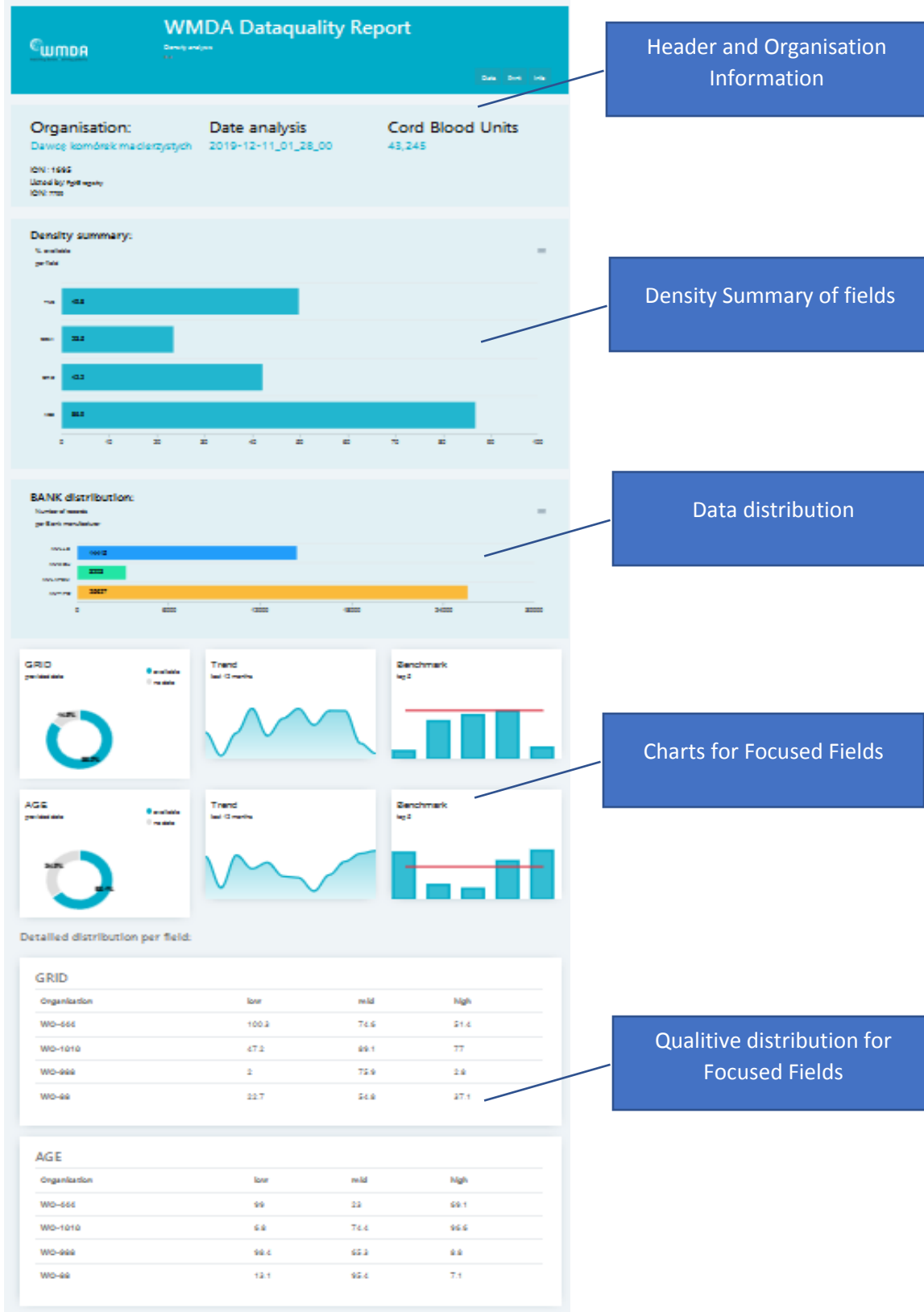
Figure 11. Example of the WMDA organisation DQ report.
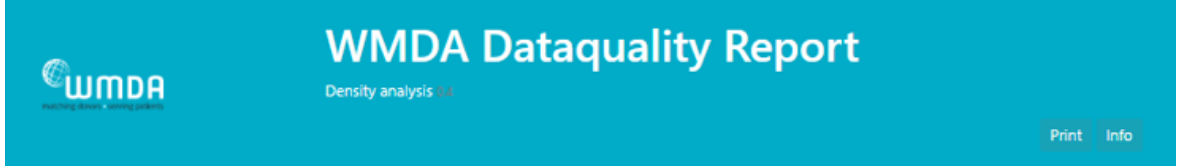*For clarity in this case, WMDA limited the data to 2 fields and used a dummy organisation*
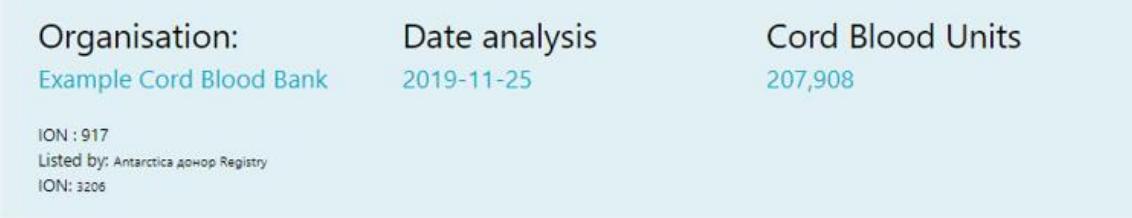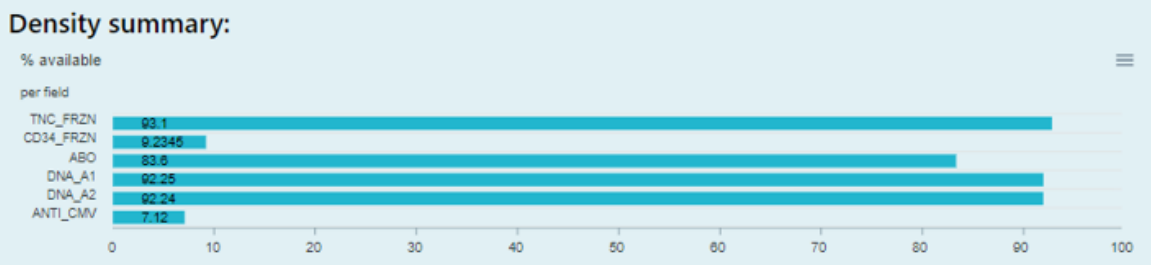
2.7.2 Report explanation

There are 5 main parts in the report as in *Figure 11*:

- o **Header and Organisation Information:** provides the report general info like name and version, and the necessary information for each organisation who has an ION and list data in WMDA international Donor / CBU database.
- o **Density Summary of Fields:** provides density report in percentage (%) for all the fields that are not empty.
- o **Data Distribution**: provides the overview of the records distribution of the CBBs affiliated with the listing CBB.
- o **Charts for Focused Fields**: provides detail chart report for the fields that WMDA focused to work on to improve the density in 2019 and 2020. For each field, 3 charts are provided, see details in the *Table 5*.
- o **Qualitive Distribution for Focused Field**: provides the value distribution of the fields that WMDA focused to work to improve the density in 2019 and 2020.

Table 5. Detailed explanation of different part of the WMDA DQ report

Section	Report example and Explanation
Header	The report's name and version. Extra feature like print report and link for information to DQ registry report is provided.
Header	Identifying organisation's name and ION, listing organisation name and ION if applicable, date that the report is generated.
Data Summary	The subdivision within the dataset for the listing organisation, typically per bank that included in the listing bank, or per ION that data uploaded by other listing organisations uploaded.
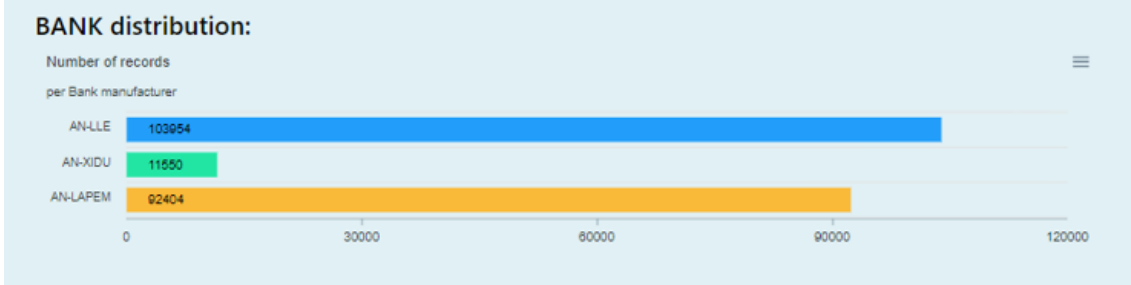
Data Distribution	**BANK distribution:** Number of records per Bank manufacturer AN-LLE — 103954 AN-XIDU — 11550 AN-LAPEM — 92404 (axis: 0, 30000, 60000, 90000, 120000)
	This bar chart provides the data distribution of records number in the organisation when the organisation is in charge of and uploading data for other CBBs and registries.
Charts for Focused fields	**VOL_FRZN** provided data ● available ○ no data 6.5% 93.5%
	This pie chart provides a visual indication of the density of the what % of the data is available for that field.
Charts for Focused fields	**Trend** last 12 months
	This line chart provides a trend analysis with the development of the % available data for that field for the given period. The trend is displayed RELATIVE. The global trend (the number of records provided over a period) can be found here: https://statistics.wmda.info/spark/.
Charts for Focused fields	**Benchmark** top 5
	This combo chart provides a comparison of the density (%) available data from the top 5 global performers that in WMDA international database. You For the global development of the AVERAGE availability of data you may consult the global density trend analysis per field at https://statistics.wmda.info/density/
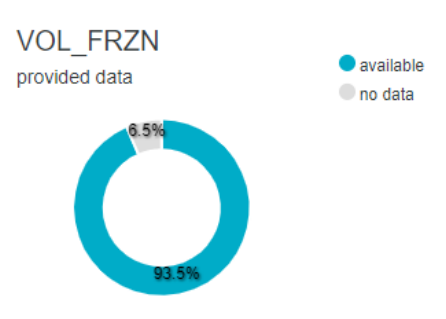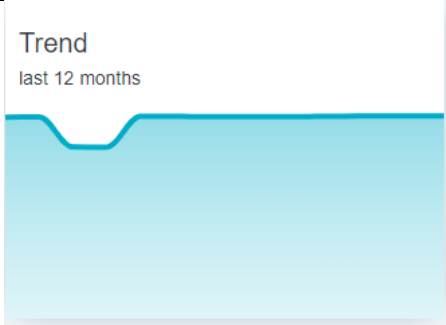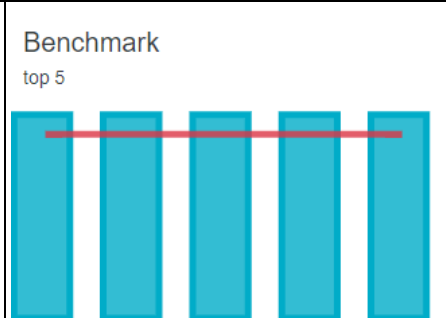
Qualitive Distribution for Focused Field	

TNC_FRZN

Organisation	[1-49]	[50-89]	[90-124]	[125-149]	[150-199]	[200-249]	[250-299]	[300-699]	[700-3000]	Missing
1111	0	3	215	472	614	223	64	34	0	0
2222	316	2653	5637	4239	4921	1258	338	127	1	0
3333	230	4403	9029	5030	4151	933	245	50	1	5
4444	303	2492	1774	697	542	127	32	11	0	14
5555	61	1618	3603	1437	947	142	19	6	0	9
Missing	16	18	10	1	3	2	2	0	4	1

This table provide the CBU count distribution with value range of the specific field. The data ranges like "[90-124]" are the default data ranges WMDA uses based on clinical requirement or community suggestions. Organisation column shows the unique WMDA organisation id (WO-ID). "Missing" in last row means some CBUs are provided without the WO-ID. "Missing" in the last column means no value provided for this field for that number of CBUs.

2.7.3 Suggestions for Organisations

In the report, some fields are with text which gives suggestions of data improvement. These suggestions are based on general recommended data request of hematopoietic stem cell transplantation (HSCT) or CBU transplantation.

2.7.4 Improvement plan

The report is missing useful features or possibility for customization, so the following improvements are considered for the next updated version:

- HLA Low and High resolution
- HLA diversity
- Improve the benchmark (not just top 5 registries with 100%)
- Break down by GRID for donor
- Customize report for organisations if needed or requested

2.8 XSD 2.2

XSD schema is a single truth file to be used for XML file validation across all platforms, including: EMDIS, WMDA Search & Match and WMDA forms. The schema may change time by time based on Request For Changes (RFCs) of data fields, that recommendations from WMDA Data Dictionary (DD) working group. These changes mainly focus on data accuracy, introducing of new data.

The benefits of using XSD are as followed:
- Help to validate one single XML structure is used for inventory (donor or CBU) data exchange
- Provide general validation rules agreed by the community and approved by DD group
- Flexible and easy to control and deploy the changes for XML

After the first release of XSD 2.1 in 2017, there was no maintenance plan. DD technique group - XSD group - was created after being discussed in the DD group, and then started to work on the plan and standard for   XSD   update. Based   on   the recommended RFCs   from the   DD group and the remaining improvements request of the XSD from 2017, the XSD group prepared a new XSD version 2.2,   which   will   be   the   new XML   schema for   the   WMDA   data   upload service. The updated schema was published   on 30   Augustus   2019 for   public   comments and   became final and was released on 1 October 2019.   The detailed changes and comparisons to the current XSD 2.1 are published here.

WMDA   implemented   XSD   2.2   in November 2019 and   will   released   this   feature on   16   December 2019. Considering the relatively slow transition time from the organisation side to use a new XML file comply with new XSD version, WMDA comes up to some decisions for a long term XSD support as below:
- XSD main version like 2.1 and 2.2, will be updated yearly, Q4 is preferred
- Small changes that have no effect to the structure are possible to be arranged quarterly with version number like 2.2.1, 2.2.2
- WMDA data upload system will always support 2 XSD versions
- XSD transition period is one year for organisations

3. Project planning for 2020

Several DQ projects in 2019 still need further attention in 2020. They are listed below:

Sub-projects	Description	Status
Density Improvement plan	Increase availability of data	Next to continuous monitoring the focused fields in 2019, WMDA will include new fields that need attention: • CCR5 • HLA high resolution typing • CT_COMPLETE_DATE
Dashboards : Data upload Statistic dashboard Density report	Provide insight in qualitative and quantitative aspects of the delivered data	• Include HLA high resolution typing results in the dashboards • Include HLA diversity analysis
Differential upload	Mechanism to push changes instead of full datasets	Design, implementation and public user guidance are done. Further rollout is planned in Q1, 2020
XSD 2.3	Agreed formal and technical description of the datasets, both for data upload to WMDA as mutual exchange	Currently, there are several RFCs under discussion in the WMDA DD committee. These changes are considered to be supported in XSD 2.3. As planned, it will be released in Q4, 2020
Online Data Dictionary	Human readable dictionary of data elements and representation in various domains	The datadictionary will be updated to reflect the changes in XSD 2.2

New DQ sub-projects that are planned for 2020

Sub-projects	Description	Data Characteristics	Status
Database Refactoring of Data Upload	System performance improvement	Timeliness	Investigation already done Arranged in Q1, 2020
Side-load projects	Improve the data load frequently from once a day to real time in Search & Match	Timeliness	Investigation already done Arranged in Q2, 2020

Appendix 1

Full density EU VS Non-EU report

Density (data available) in %				
	Non-EU	EU member states	Non-EU	EU member states
Data element	Cord Blood Units	Cord Blood Units	Donors	Donors
A1	24.1	23.2	22.9	35.2
A2	22.2	20.9	20.6	30.6
ABO	69.2	73.0	44.6	51.5
AL_PLA	3.2	11.7		
AL_RED_BC	-	9.4		
AL_SER	0.1	10.2		
ALT		4.3		1.0
ANTI_HBC	7.8	10.5	0.0	1.1
ANTI_HBS	-	4.5	0.0	0.3
ANTI_HCV	9.4	10.0	0.1	3.6
ANTI_HIV_12	9.4	7.7	0.1	0.6
ANTI_HTLV	6.3	11.3	0.1	2.6
ATT_SEG	15.6	15.0		
B1	24.0	23.2	22.8	35.3
B2	23.1	22.0	21.5	32.6
BACT_CULT	25.3	20.1		
BAG_ID	9.4	14.8		
BAG_TYPE	12.5	8.9		
BAGS	12.5	28.3		
BANK_DISTRIB_ID	49.8	75.0		
BANK_MANUF_ID	60.7	75.0		
BANK_MAT_ID	3.1	4.3		
BIRTH_DATE	81.3	100.0	91.3	100.0
C1	2.5	8.2	2.3	13.1
C2	2.3	6.3	1.6	7.8
CCR5	3.1	5.8	11.7	8.6
CD34PC	3.1	21.1		
CD34PC_FRZN	68.2	80.6		
CFU_FRZN	13.9	19.6		
CHAGAS		4.8		
CHECKUP_DATE			0.5	1.1
CMV	12.0	11.5	4.6	16.2
CMV_DATE	12.0	13.9	4.6	16.6
CMV_NAT		8.5		0.3
COLL_DATE	36.7	33.0		
COLL_TYPE			5.2	12.3
CONTACT_DATE			7.1	24.0
CT_COMPLETE_DATE	3.2	4.3		
CT_SMPL_TYPE	3.2	4.3		
DNA_A1	88.2	93.5	87.2	84.5
DNA_A2	84.7	84.7	84.2	76.7
DNA_B1	88.2	93.6	87.0	84.5
DNA_B2	86.3	89.4	85.3	80.5
DNA_C1	39.2	53.0	56.1	63.1
DNA_C2	37.9	49.7	54.7	58.8
DNA_SMPL	4.2	12.3		
DON_ATTR	9.4	4.2		
DON_POOL	100.0	100.0		
DPA11	0.2	0.2	1.6	0.6
DPA12	0.1	0.1	1.4	0.4
DPB11	3.6	11.6	28.7	30.1
DPB12	3.4	10.3	27.8	26.3
DQ1	3.9	3.1	2.9	5.1
DQ2	3.9	3.0	2.8	4.2
DQA11	0.3	2.9	3.2	2.9

DQA12	0.3	2.5	3.1	2.7
DQB11	15.7	36.3	46.5	47.8
DQB12	15.4	32.4	45.3	43.8
DR1	20.5	19.7	11.0	17.6
DR2	19.8	18.5	10.5	15.8
DRB11	92.9	99.6	89.1	86.2
DRB12	89.8	93.6	86.7	80.7
DRB31	5.5	6.1	7.3	5.5
DRB32	0.8	0.7	1.6	0.9
DRB41	3.3	4.5	5.6	4.7
DRB42	0.3	0.1	0.5	0.1
DRB51	2.7	5.2	4.0	2.7
DRB52	0.1	0.0	0.3	0.1
EBV	-	4.3	0.0	0.2
ETHN	20.8	29.7	35.4	34.0
FREEZE_DATE	15.8	19.5		
FREEZE_METH	9.6	19.5		
FUNG_CULT	25.3	20.1		
GRID			89.0	94.3
HBS_AG	9.4	9.8	0.1	3.5
HBV_NAT	6.3	7.5	0.0	0.1
HCV_NAT	6.3	12.2	0.0	0.1
HEIGHT			12.6	14.7
HEMO_STATUS	18.7	11.1		
HIV_1_NAT	6.3	11.9	0.0	0.1
HIV_P24	3.1	6.6		1.2
KIR_GL		-		
KIR2DL1	-	4.2	5.2	4.6
KIR2DL2	-	4.2	5.2	4.6
KIR2DL3	-	4.2	5.2	4.6
KIR2DL4	-	4.2	5.2	4.6
KIR2DL5A	-	4.2	5.2	4.6
KIR2DL5B	-	4.2	5.2	4.6
KIR2DP1	-	4.2	1.0	4.6
KIR2DS1	-	4.2	5.2	4.6
KIR2DS2	-	4.2	5.2	4.6
KIR2DS3	-	4.2	5.2	4.6
KIR2DS4	-	4.2	5.2	4.6
KIR2DS5	-	4.2	5.2	4.6
KIR3DL1	-	4.2	5.2	4.6
KIR3DL2	-	4.2	5.2	4.6
KIR3DL3	-	4.2	5.2	4.6
KIR3DP1	-	4.2	5.2	4.5
KIR3DS1	-	4.2	5.2	4.6
LOCAL_ID	13.3	50.0		
MAT_A1	-	2.7		
MAT_A2	-	2.6		
MAT_AL_PLA	3.2	10.9		
MAT_AL_SER	0.1	11.3		
MAT_ALT	-	4.8		
MAT_ANTI_HBC	12.5	17.7		
MAT_ANTI_HBS	0.0	4.9		
MAT_ANTI_HCV	18.3	19.7		
MAT_ANTI_HIV_12	13.9	20.1		
MAT_ANTI_HTLV	15.1	16.3		
MAT_B1	-	2.8		
MAT_B2	-	2.7		
MAT_C1	-	0.1		
MAT_C2	-	0.1		
MAT_CHAGAS	6.1	4.8		
MAT_CMV	19.3	18.3		
MAT_CMV_DATE	10.5	7.0		
MAT_CMV_NAT	-	4.2		

MAT_DNA_A1	3.7	10.5		
MAT_DNA_A2	3.5	10.2		
MAT_DNA_B1	3.7	10.4		
MAT_DNA_B2	3.6	10.2		
MAT_DNA_C1	3.7	5.2		
MAT_DNA_C2	3.5	5.1		
MAT_DPA11	-	-		
MAT_DPA12	-	-		
MAT_DPB11	-	2.5		
MAT_DPB12	-	2.5		
MAT_DQ1	-	0.2		
MAT_DQ2	-	0.2		
MAT_DQA11	-	-		
MAT_DQA12	-	-		
MAT_DQB11	0.0	5.1		
MAT_DQB12	0.0	5.0		
MAT_DR1	-	0.5		
MAT_DR2	-	0.5		
MAT_DRB11	3.7	8.2		
MAT_DRB12	3.5	8.0		
MAT_DRB31	-	-		
MAT_DRB32	-	-		
MAT_DRB41	-	-		
MAT_DRB42	-	-		
MAT_DRB51	-	-		
MAT_DRB52	-	-		
MAT_EBV	-	8.6		
MAT_HBS_AG	12.5	22.6		
MAT_HBV_NAT	9.4	13.9		
MAT_HCV_NAT	11.8	15.6		
MAT_HIV_1_NAT	11.8	14.5		
MAT_HIV_P24	3.7	10.9		
MAT_ID	9.4	11.1		
MAT_PB19_NAT	-	4.2		
MAT_PLA_QUANT	3.1	14.8		
MAT_SER_QUANT	1.1	15.0		
MAT_SYPHILIS	15.2	21.7		
MAT_TOXO	0.0	9.5		
MAT_WNV	6.6	4.3		
MNC_FRZN	16.2	34.4		
NMBR_MARR			28.3	51.5
NMBR_PBSC			28.3	51.5
NMBR_PREG			6.4	11.6
NMBR_TRANS			14.6	35.3
OTH_SMPL	6.3	5.8		
PB19_NAT		6.8		0.0
PLA_QUANT	3.2	14.7		
PROC_DATE	15.5	25.4		
PROC_METH	11.1	30.4		
PROC_METH_TYPE	10.5	22.3		
PROD_MOD	15.7	10.8		
RED_BC_FRZN	5.4	9.6		
RHESUS	66.2	73.0	41.6	51.1
RSV_PAT		2.1		
SER_QUANT	-	14.3		
SEX	60.8	90.1	93.5	100.0
STAT_END_DATE	-	2.2	0.0	0.2
STAT_REASON		0.1		0.3
STATUS	100.0	100.0	100.0	100.0
SYPHILIS	9.4	11.2	0.1	3.4
TNC	15.3	30.7		
TNC_FRZN	98.1	94.3		
TOXO	-	4.8	0.0	0.5

VIABILITY	16.7	19.7		
VIABILITY_CELLS	14.7	12.6		
VIABILITY_DATE	8.4	13.8		
VIABILITY_METHOD	19.7	12.5		
VOL	40.0	76.4		
VOL_FRZN	88.4	90.6		
WEIGHT			14.8	23.6
WNV		4.2		