

Summary on the alignment of ENVRI-RM and RDA DFT WG outputs.

The DFT WG has completed its initial cycle and it is ready to move into an IG. The outputs of the DFT WG are described in Table 1. The main output of the group is contained in the Results RFC [6] report which is the current snapshot and contains the terms and their relationships. This document is provided with the suffix RFC, meaning that is in its final phase for review by the community and to become a standard. The document contains 14 terms, their definitions and relationships. The document also presents a diagram describing the relationships of the terms.

**Table 1 DFT WG Summary of Outputs**

<b>Title</b>	<b>Description</b>	<b>Comment</b>	<b>Ref.</b>
RDA Data Foundation and Terminology DFT 1: Overview	Presents 22 Data Models which are being considered for building a common model	Yin's model is discussed	4
RDA Data Foundation and Terminology DFT 2: Analysis & Synthesis	Presents and analysis and synthesis of the models described in the first document	Presents the first version of the data model.	7
RDA Data Foundation and Terminology DFT 3: Snapshot of DFT Core Terms	Presents the first definition of core terms.	Highlights this as a snapshot of work in progress towards real, working agreements on terminology within RDA and across the worldwide data community	8
RDA Data Foundation and Terminology DFT 4: Use Cases	Presents a series of use cases and how they use the core terms to describe them.	Assert that EUDAT is largely aligned and influenced by RDA work Describes CLARIN as an early DFT adopter Talks about data streams but calls them "gappy dynamic data"	5
RDA Data Foundation and Terminology DFT 3: Term Tool Description <sup>1</sup>	Presents an online tool for collecting terms provided by the adopting community	Online tool contains 159 term definitions, interesting to see what they are.	1
RDA Data Foundation and Terminology - DFT: Results RFC	First RFC on the terms identified by the DFT	This contains the definitions of 14 concepts and their relationships. This is a basic ontology for classifying data objects and repositories	6
DFT WG Products Metadata	Metadata about documents		1
Data Foundation and Terminology Work Group Products Maintenance and Retirement plan.	How to pass on the product from the WG to an IG	The retirement plan contemplates the transition to the DFT IG which should continue the development of the model	9
RDA Data Foundation and Terminology DFT: Adoption Note		Describes the flyer as the main entry point in addition to the other documents. Points to interaction with EUDAT and ESFRI RIs as key for further development	3

<sup>1</sup> Error the title says 3 and it should be 5, according to the outline included in most of the DFT WG documents.

## ENVRI-RM Alignment to DFT Core Term Definitions

The terms defined in the ENVRI-RM information viewpoint (IV) are the closest match to those described by the DFT WG. Table 2 includes a comparison of the terms in the IV and then the corresponding term in the DFT Core Term Definitions (DFTWG-CTD).

Table 2 IV terms compared to DFT WG terms

IV Term	DFTWG-CTD	Reasoning
specification of investigation design	Metadata	Describes the reason for collecting the digital object
specification of measurements or observations	Metadata	Describes the method for collecting the digital object
measurement result	Bitstream	Direct mapping, in this case the bit stream is the digitalised form of a reading. How do we deal with physical and analogue recordings?
concept	Metadata	Assigns meaning and links it to other metadata or to a digital object
conceptual model	Digital Collection	Is an aggregation of concepts, thus an aggregation of metadata
QA notation	Digital Collection	Is an aggregation of concepts related to the quality of data and metadata
metadata	Metadata	Direct link but the meaning given by DFTWG is more strict and brief
metadata state	state information	Describes state of metadata
metadata catalogue	Digital Collection	Aggregation of digital objects
citation	Persistent Identifier	Long lasting ID that links to a digital object
persistent data	Digital Object	Direct link
data state	state information	Describes state of digital object
unique identifier (UID)	Persistent Identifier	Long lasting ID that links to a digital object
backup	Digital Object	
mapping rule	Metadata	Describes allowed operations on data
data provenance	Metadata	Describes the digital object
service		No correspondence directly but PID resolver conversely could be seen as a service
service description	Metadata	Describes the digital object

The DFTWG definitions and the model are so simple that it is not hard to map it to concepts in the IV. Using the DFTWG then leaves the IV terms as a set of data and metadata terms. Should be interesting to see the relationships amongst the terms in the IV (is this Paul's work).

Table 3 compares the terms from the DFTWG to terms in the ENVRI-RM

**Table 3 DFT WG terms compared to IV Terms**

<b>DFTWG-CTD</b>	<b>IV Term(s)</b>	<b>Reasoning</b>
Digital Object	All information objects, with the exception of service, can be characterised as digital objects	The definitions of information object and metadata allow for this
Persistent Identifier (PID)	unique identifier (UID)	Long lasting ID that links to a digital object
PID Record	metadata, data provenance	The PID record is a set of attributes describing a Digital Object
PID Resolver	Service	A service (or system) globally available for resolving a PID
Metadata	specification of investigation design, specification of measurement or observation, conceptual model, QA notation, metadata, metadata catalogue citation, concept, data state, metadata state, back up, data provenance, mapping rule, service description.	All metadata and collections in the IV can be seen as types of metadata
Aggregation	specification of investigation design, specification of measurement or observation, conceptual model, QA notation, metadata catalogue citation, concept, data state, metadata state, back up, data provenance, service description.	All the collections in the IV are aggregations of Digital Objects
Digital Collection	specification of investigation design, specification of measurement or observation, conceptual model, QA notation, metadata catalogue citation, concept, data state, metadata state, back up, data provenance, service description.	All the aggregations are digital collections as well
Digital Entity	All information objects can be characterised as digital entities	By definition
Repository	Service	By definition
Bitstream	persistent data	By definition
State information	Data state, metadata state	By definition
Property	Metadata	By definition
Metadata Repository	Metadata catalogue, service	By definition
Checksum	QA Notation, Metadata	By definition

The IV does not have a definition of Semantic Annotation. Neither does the DFTWG-CTD.

## References

1. DFT WG – RDA (2014). DFT WG Products Metadata. Creator: Gary Berg -Cross, Raphael Ritz, Peter Wittenburg. Date 12/10/2014. Consulted on 04/03/2016. Available at: <https://rd-alliance.org/system/files/DFT%20WG%20%20Products%20metadata.pdf>

2. DFT WG – RDA (2014). RDA Data Foundation and Terminology DFT 3: Term Tool Description. Eds. Thomas Zastrow, Gary Berg-Cross, Raphael Ritz. Date: 07/12/2016. Consulted on: 04/03/2016. Available at: <https://rd-alliance.org/system/files/DFT%20term%20tool-dec-1-0.pdf>
3. DFT WG – RDA (2014). RDA Data Foundation and Terminology DFT: Adoption Note. Eds. Gary Berg -Cross, Raphael Ritz, Peter Wittenburg. Date 31/12/2014. Consulted on 04/03/2016. Available at: <https://rd-alliance.org/system/files/adoption-notes-1-2.pdf>
4. DFT WG – RDA (2014). RDA Data Foundation and Terminology DFT 1: Overview. Date 31/12/2014. Eds. Gary Berg-Cross, Karen Green, Peter Wittenburg. Consulted on 04/03/2016. Available at: <https://rd-alliance.org/system/files/RDA%20DFT%20Data%20Models-v1-6.pdf>
5. DFT WG – RDA (2014). RDA Data Foundation and Terminology DFT 4: Use Cases. Date 31/12/2014. Eds. Gary Berg, Karen Green, Peter Wittenburg. Consulted on 04/03/2016. Available at: <https://rd-alliance.org/system/files/Use%20case%20v1%206.pdf>
6. DFT WG – RDA (2015). RDA Data Foundation and Terminology - DFT: Results RFC. Eds. Gary Berg -Cross, Raphael Ritz, Peter Wittenburg. Date: 29/06/2015. Consulted on: 04/03/2016. Available at: <https://rd-alliance.org/system/files/DFT%20Core%20Terms-and%20model-v1-6.pdf>
7. DFT WG – RDA (2015). RDA Data Foundation and Terminology DFT 2: Analysis & Synthesis. Eds. Gary Berg-Cross, Karen Green, Raphael Ritz, Peter Wittenburg. Date: 31/07/2015. Consulted on: 04/03/2016. Available at: <https://rd-alliance.org/system/files/RDA%20DFT%20Data%20Models-analysis-v1-5.pdf>
8. DFT WG – RDA (2016). RDA Data Foundation and Terminology DFT 3: Snapshot of DFT Core Terms. Eds. Gary Berg-Cross, Keith Jeffery, Bob Kahn, Larry Lannom, Raphael Ritz, Herman Stehouwer, Peter Wittenburg, Thomas Zastrow, Zhu Yunqiang. Date: 31/07/2015. Consulted on: 04/03/2016. Available at: <https://rd-alliance.org/system/files/DFT%20Core%20Terms-dec-1-5.pdf>
9. DFT WG – RDA (2016). Data Foundation and Terminology Work Group Products Maintenance and Retirement plan. Date: Not available assumed 2016. Consulted on: 04/03/2016. Available at: <https://rd-alliance.org/system/files/Data%20Foundation%20and%20Terminology%20Work%20Group%20Products%20Maintenance%20and%20Retirement%20plan.pdf>