The DAF DDI Profile, a Metadata Set to Address Digital Curation and Preservation Issues in Cultural Heritage Institutions

Ruggero Lancia

Humanities Advanced Technology and Information Institute (HATII),
George Service House, 11 University Gardens,
University of Glasgow, Glasgow, G12 8QH, Scotland, UK
R.Lancia@hatii.arts.gla.ac.uk

Abstract. This investigation, funded by the Digital Curation Centre (DCC) and conducted in the Humanities Advanced Technology and Information Institute (HATII) of the Glasgow University, is part of the activities within the Data Asset Framework (DAF) to provide research organisations with an online tool to identify, locate, describe and assess how they are managing their research data assets. This research has produced a metadata set fitted for the implementation of curation and preservation policies in small research institutions. In this paper, the DAF DDI profile will be exposed and explained. Further, it will be discussed, the advantages this metadata set offers to the small Cultural Heritage institutions for supporting curation and preservation policies. The DAF DDI profile is the first example of a DDI 3.1 implementation in a general curatorial context, that is not disciplinary specific, and its integration with PREMIS 2.0 is the first published mapping of PREMIS to DDI.

Keywords: DAF Interactive Tool, DCC, Metadata, DDI, PREMIS, *Digital Curation and Preservation.*

1 Introduction

This investigation, funded by the Digital Curation Centre (DCC) and conducted in the Humanities Advanced Technology and Information Institute (HATII) of the Glasgow University, is part of the activities within the Data Asset Framework (DAF) to provide research organisations with an online tool to identify, locate, describe and assess how they are managing their research data assets.

This study was aimed at developing a metadata set to ease the satisfaction of the requirements attached to the DAF interactive tool redevelopment, that is a) the possibility for the users to extend and customize its audit forms, b) the improvement of the analysis tools, such as for example implementing automatic recommendations facilities, c) and the interoperability with the DRAMBORA tools, available at http://www.repositoryaudit.eu/ (accessed 14 June 2010). In particular, the need for extensible and customizable audit forms has been interpreted here as the need for a metadata set flexible, modular and extensible, while to address the interoperability

M. Ioannides (Ed.): EuroMed 2010, LNCS 6436, pp. 291–305, 2010.

and reuse of data across the DCC framework, the researched metadata set has been expected to be interoperable and pertinent to the entire curatorial lifecycle of digital assets.

Featuring all these characteristics, the DDI 3.1 metadata set has been chosen as a basis for this study. In effect, from DDI 3.0, this metadata set supports a lifecycle model, the "Combined Lifecycle Model", including as curatorial steps the conception, creation, ingestion, storage, access and reuse. Further, the DDI 3.1 offers a modular structure made of 21 XML namespaces, which can be assembled without restrictions but maintaining the consistence of the hierarchical relationships among their schemas, and allows for external extensions as long as the general DDI profile is not altered. This characteristic makes the DDI 3.1 easily interoperable with other metadata sets, for example the DCMI elements have been integrated in the DDI documentation within a specific namespace and mappings to other sets have been already released by the DDI developers.

So, at the beginning of this study the DDI structure has been analysed evaluating its use to document the curation and preservation practices. This investigation has demonstrated a weakness of the DDI set managing the assets' preservation and its need for the integration with a metadata set specifically designed for preservation purposes. For this reason, a mapping of the PREMIS 2.0 core elements to DDI 3.1 has been developed to compensate these shortcomings: a major accomplishment of this research since it is the first mapping between these metadata set to be so far completed and published. Then, the researched metadata set has been designed as a DDI profile resulting from the selection of DDI elements accomplishing the objectives of this investigation.

In this paper, the DAF DDI profile will be exposed and the selection process of its elements will be explained giving prominence to the mapping of PREMIS to DDI. Further, it will be discussed, the advantages this metadata set offer to the small Cultural Heritage institutions for supporting curation and preservation policies.

2 DDI 3.1 as Basis for the DAF Metadata Set

Looking in this study for a metadata set modular, customizable and extensible, the DDI 3.1 set appeared to be the fittest option. That assumption was comforted by its wide adoption in prestigious institutional archives; nevertheless, because of its disciplinary specialisation, this metadata set was expected to cause difficulties being implemented in a general framework such as the one of the DAF methods and tools.

Thus, in this study, the implementation of the DDI 3.1 as basis for the researched metadata set has been assessed against its shortcomings serving the curation and preservation purposes of a repository from a research institution.

2.1 DDI Shortcomings Serving the Curation and Preservation Purposes

The DDI set is claimed by its creators to support the productive lifecycle of the digital assets rather than its archival part and, in effect, from the "Combined Life Cycle Model" proposed in the "Overview" a lack of attention for all the curatorial processes but the access and reuse of assets is evident: the "Data Archiving" step in this lifecycle

model is just accessed through an alternative path on the way from “Data Processing” to “Data Distribution” and in this way the management duties of repositories are completely obliterated [1].

Thus, the DCC Digital Curation Lifecycle model, because of its general and holistic approach to curation [2], has been chosen as methodological background to analyse the DDI shortcomings recording a complete curatorial lifecycle disciplinary unspecific.

In more detail, the classification of the metadata has been limited to the sequential actions of the DCC lifecycle model, that is the curatorial steps “Create and Receive”, “Appraise and Select”, “Ingest”, “Preservation Action”, “Store”, “Access Use and Reuse” and “Transform” since the full time and the occasional ones appeared to be respectively too general and too specific for the purposes of this study.

From this analysis emerged, firstly, a bias toward few specific functions and events affecting the digital assets resulting in the insufficiency of descriptions for general curation purposes in the “Ingest”, “Store”, “Access Use and Reuse” and “Transform” sections, and second, the lack of concentration on preservation issues.

In effect, while many DDI descriptors allow to record, aiming at their reuse, the structure of statistical data and to reference their data structures and collection procedures, the element <LifecycleInformation> is the only descriptor available to describe the relevant events of the curatorial steps “Ingest”, “Store” and “Preservation action”. In a similar fashion, DDI offers just <FundingInformation> and <LifecycleInformation> to record the appraisal and the disposal of the assets in an unspecific curatorial context while a more accurate technical appraisal of the digital objects' quality is possible just for statistical data through the descriptors of <ProcessingEvent> and <CollectionEvent>. Similarly, at the “Access Use and Reuse” step, the DDI set offers many administrative and structural metadata especially related to quantitative analysis for social sciences to manage accessibility and reuse of the assets but no descriptors to record their reuse apart from referencing elements such as <source>, <OtherMaterial>and <Comparison>.

On the preservation's side, the descriptors of <LifecycleInformation> allow a listing of events in the lifecycle of a data set or collection but do not offer specific descriptors to facilitate the automation of this kind of report, nor the sufficient guide for a human being to relate his description to preservation levels and curatorial policies, that is to assess his activity on the repository and to scale his efforts according to feasibility considerations.

2.2 The DDI 3.1 as a Basis for the Researched Metadata Set

The disciplinary specialisation of this metadata set is evident also considering its structure.

Offering 6 specific namespaces to allow the computational reuse of quantitative social sciences data, the structure of the DDI metadata reflects its main application's field. In effect,

“ddi:physicaldataproduct_ncube_normal:3_1”,
“ddi:physicaldataproduct_ncube_inline:3_1”,
“ddi:physicaldataproduct_ncube_tabular:3_1”,

"ddi:physicaldataproduct_proprietary:3_1", "ddi:logicalproduct:3_1" and "ddi:datacollection:3_1" describe coding and variables of the assets supporting verification and reuse of quantitative researches.

Similarly, this set of metadata aims at being interoperable with other metadata standards such as Dublin Core, MARC, ISO11179 (metadata registries), SDMX (data exchange), and geographic standards such as FGDC (Federal Geographic Data Committee) and ISO 19115 which are widely diffuse in the social sciences archives.

Nevertheless, the DDI set's interoperability depends on its modularity, which allows also the creation of metadata registries, the import of external namespaces and the grouping and versioning of both data and documentation. Further, the modular approach of this metadata is due to the adoption of a data lifecycle model which, although defective dealing with the preservation management, offers these advantages for a curatorial framework: a) to allow the capture and preservation of metadata generated by different agents at different points in time; b) to track changes and updates in both data and documentation; c) to enable investigators, data collectors and producers to document their work directly in DDI; d) to benefit data users who need information from the full data lifecycle for optional discovery, evaluation, interpretation and reuse of data resources [3].

In this way, the DDI 3.1 allows also a strong customization adopting its descriptors through the identification of a profile, that is a selection of descriptors common to a collection or group, and at the same time enough flexibility to render compliant also incomplete records, supporting the feasibility of its implementation. Moreover, the DDI content can be integrated in digital repository systems, such as DSpace, e-Prints and Fedora, into data search system allowing queries across multiple modules, and into data search and manipulation tools, in doing so fostering the implementation of automated curatorial procedures.

For these reasons, considering the benefits offered by DDI complementing its shortcomings, DDI 3.1 has been chosen in this research as basis for the researched metadata, which has been designed as a DDI profile.

3 Integrating PREMIS in DDI

Since the researched metadata set has been considered mainly intended for small research institutions in a very early stage of their curatorial procedures' implementation, the DDI profile resulting from this research was expected to be a light metadata model aiming less at the completeness of the recorded information than at the feasibility of its implementation. Thus, despite the modular structure of DDI allows external extensions, a mapping to other metadata sets with a specific curatorial interest has been considered the fittest option to obtain a DDI profile retaining the adequate information to support curatorial procedures in small contexts. Otherwise, the extension would had caused expansive redundancies of the information that if resolved, for example selecting the metadata elements to be imported as DDI extension in a separate namespace, would had caused partial validations of the metadata records.

In particular, since the release of DDI 3.0, the PREMIS standard has been reported to have been analysed for mapping to the DDI model in order to face the preservation issues arising in research repositories [4]. Encouraged by this previous attempt, in this research it has been decided to integrate the PREMIS semantic units in the DAF DDI profile through a mapping.

3.1 Difficulties Integrating PREMIS Semantic Units in DDI

The structural differences between PREMIS and DDI cause many difficulties mapping the one into the other. In effect, the linking among the five core entities proposed by the PREMIS model, that is "Intellectual Entities", "Objects", "Events", "Rights" and "Agents", compared to the horizontal description offered by DDI, implies a considerable loss of information in the translation.

Further, despite the stated limit of the PREMIS 2.0 to offer descriptive metadata, the granularity this set offers by individuating the entities' "typologies" as "representations", "files" and "bitstreams" is not repeated in DDI 3.1 (PREMIS, 2008). As suggested in a draft of the mapping of PREMIS 1.0 to DDI 3.0 released in December 2007 in the ICPSP online forum [5], the asymmetry between those metadata set is best controlled by considering the entire information on the assets related just to the "file" typology.

Similarly, also significant information for the long-term preservation of the assets is inevitably lost through the integration of PREMIS in DDI, that is the semantic units "1.3 preservationLevel", "1.5.1 compositionLevel" (especially "1.5.2 fixity", "1.5.3 size" and "1.5.6 inhibitors"), "1.8 environment" (a part from its sub-unit "1.8.4 dependency") and "1.9 signatureInformation". In effect, these descriptors record respectively the information indicating the policy on the set of preservation functions to be applied to an object, the context in which the policy was made, the technological environment supporting the use of the object and the authoritative quality of the assets.

Because of the limited preservation policies small research institutions are likely to apply, this kind of information's incompleteness concerning the digital assets' preservation has been considered tolerable for the researched DDI profile. For this reason, the selection of the PREMIS semantic units to be mapped to DDI has been narrowed down to the elements declared mandatory by the PREMIS XML schema.

The vast majority of the information requested by this schema concerns the entity "Objects" and is repeated, if applicable, for the files, the files compound and the bitstreams in order to obtain a granular curatorial description of the intellectual entities' components. The basic description of each file includes its identifier and its format through the elements <format> of <objectCharacteristics> and <objectIdentifier>, as it could be noticed reading the relative selection of units from the PREMIS XML schema featured below, see http://www.loc.gov/standards/premis/premis.xsd (accessed 14 June 2010):

```
<xs:complexType name="file">
<xs:complexContent>
<xs:extension base="objectComplexType">
<xs:sequence>
<xs:element ref="objectIdentifier" minOccurs="1"
maxOccurs="unbounded"/>
<xs:element ref="objectCharacteristics" minOccurs="1"
maxOccurs="unbounded"/>
</xs:sequence>
<xs:attribute name="xmlID" type="xs:ID"/>
</xs:extension>
</xs:complexContent>
</xs:complexType>
```

```
<xs:complexType name="objectIdentifierComplexType">
<xs:sequence>
<xs:element ref="objectIdentifierType" minOccurs="1"
maxOccurs="1"/>
<xs:element ref="objectIdentifierValue" minOccurs="1"
maxOccurs="1"/>
</xs:sequence>
<xs:attributeGroup ref="xlink:simpleLink"/>
</xs:complexType>
<xs:complexType name="objectCharacteristicsComplexType">
<xs:sequence>
<xs:element ref="format" minOccurs="1"
maxOccurs="unbounded"/>
</xs:sequence>
</xs:complexType>
<xs:complexType name="formatComplexType">
<xs:sequence>
<xs:choice>
<xs:sequence>
<xs:element ref="formatDesignation" minOccurs="1"
maxOccurs="1"/>
</xs:sequence>
<xs:element ref="formatRegistry" minOccurs="1"
maxOccurs="1"/>
</xs:choice>
</xs:sequence>
</xs:complexType>
<xs:complexType name="formatDesignationComplexType">
<xs:sequence>
<xs:element ref="formatName" minOccurs="1"
maxOccurs="1"/>
</xs:sequence>
</xs:complexType>
<xs:complexType name="formatRegistryComplexType">
<xs:sequence>
<xs:element ref="formatRegistryName" minOccurs="1"
maxOccurs="1"/>
<xs:element ref="formatRegistryKey" minOccurs="1"
maxOccurs="1"/>
</xs:sequence>
<xs:attributeGroup ref="xlink:simpleLink"/>
</xs:complexType>
```

Although in the PREMIS 2.0 data dictionary <storage> is listed erroneously as mandatory, in the PREMIS XML schema this element is not considered fundamental, thus, it has not been mentioned here. The absence of the mandatory element <compositionLevel> is motivated by the assumption that since in this mapping the digital objects are described just at the file level there is little need to indicate if the object is subject to processes of decoding; indeed, its default value, 0, indicates that the object described is a base object not subject to further decoding, and that could really be more often the case of the assets the DAF prototype is intended to deal with.

The “Rights” entity has a slightly less complex description concentrated on the copyright, licenses, and statute description of each object aimed at determining whether a repository has the right to perform a certain action on its assets in an automated fashion:

```
<xs:complexType name="rightsComplexType">
<xs:choice minOccurs="1" maxOccurs="unbounded">
<xs:element ref="rightsStatement"/>
</xs:choice>
<xs:attribute name="xmlID" type="xs:ID"/>
</xs:complexType>
<xs:complexType name="rightsStatementComplexType">
<xs:sequence>
<xs:element ref="rightsStatementIdentifier" minOccurs="1"
maxOccurs="1"/>
<xs:element ref="rightsBasis" minOccurs="1"
maxOccurs="1"/>
</xs:sequence>
</xs:complexType>
<xs:complexType
name="rightsStatementIdentifierComplexType">
<xs:sequence>
<xs:element ref="rightsStatementIdentifierType"
minOccurs="1" maxOccurs="1"/>
<xs:element ref="rightsStatementIdentifierValue"
minOccurs="1" maxOccurs="1"/>
</xs:sequence>
<xs:attributeGroup ref="xlink:simpleLink"/>
</xs:complexType>
```

The “Agent” and the “Event” entities have structures far more simple than the previous entities: the first offers just one mandatory element, <agentIdentifier>, while the second adopts the elements <eventType>, <eventDateTime> and <eventIdentifier> as minimum describers:

```
<xs:complexType name="agentComplexType">
<xs:sequence>
<xs:element ref="agentIdentifier" minOccurs="1"
maxOccurs="unbounded"/>
</xs:sequence>
<xs:attribute name="xmlID" type="xs:ID"/>
</xs:complexType>

<xs:complexType name="eventComplexType">
<xs:sequence>
<xs:element ref="eventIdentifier" minOccurs="1"
maxOccurs="1"/>
<xs:element ref="eventType" minOccurs="1" maxOccurs="1"/>
<xs:element ref="eventDateTime" minOccurs="1"
maxOccurs="1"/>
</xs:sequence>
<xs:attribute name="xmlID" type="xs:ID"/>
</xs:complexType>
```

3.2 Cross-Walk from PREMIS Core Elements to DDI

The following table (Table 1) summarise the proposed mapping of the PREMIS mandatory elements to DDI:

Table 1. Cross-walk from PREMIS core elements to DDI

PREMIS 2.0 core semantic unit Name	DDI 3.1 – Complex Element	DDI 3.1 – Element/ attribute	Value
(1.1) objectIdentifier	<DDI Instance>	@id	xs: string
(1.1.1) objectIdentifierType			
(1.1.2) objectIdentifierValue			
(1.5) object Characteristics			
(1.5.4) format	<Item>	<Format>	xs: string
(1.5.4.1) format Designation			
(1.5.4.1.2) formatName			
(1.5.4.2) formatRegistry	X	X	X
(1.5.4.2.1) formatRegistryName	X	X	X
(1.5.4.2.2) formatRegistryKey	X	X	X
(2.1) eventIdentifier	<Lifecycle Event>	@id	xs: string
(2.1.1) eventIdentifierType			
(2.1.2) eventIdentifierValue			
(2.2) eventType	<Lifecycle Event>	<EventType>	xs: string
(2.3) eventDateTime	<Lifecycle Event>	<Date>	xs: string
(3.1) agentIdentifier	<Agency Organization Reference>	<ID> or <URN>	xs: string or URN
(3.1.1) agentIdentifierType			
(3.1.2) agentIdentifierValue			

Table 1. (*continued*)

(4.1) rightsStatement			
(4.1.1) rightsStatementIdentifier	<Citation> and <DCElements>	<Copyright> and <rights>, respectively	xs: string
(4.1.1.1) rightsStatementIdentifier Type			
(4.1.1.2) rightsStatementIdentifier Value			
(4.1.2) rightsBasis			

But, the mapped elements are not structurally equivalent, as it is particularly evident considering the elements describing the file formats and the rights statement associated to the instances.

In effect, PREMIS, taking into account the use of extensions to apply descriptive, structural and administrative metadata, adopts a system of external references and pertinent descriptors for each of its five core entities. In this way, the "formatRegistry" unit identifies the format of an instance by reference to an entry in a format registry and the "rightsStatementIdentifier" locates the rights statement within a preservation repository system. Nevertheless, these units are options of a mandatory selection and their partners, "formatDesignation" and "rightsBasis", can be effectively expressed through the values of <Format> and <rights>. In particular, this latter would supplement the information recorded by the entity <Copyright> that might be used to express thoroughly the rights statement that would be usually referenced by "rightsStatementIdentifier".

Similarly, the units "objectIdentifier" and "eventIdentifier" can be mapped to the identifier attributes of the entities <DDIInstance> and <LifecycleEvent> despite the flattening of the information in the passage from PREMIS to DDI. The unit "agentIdentifier" can be mapped to the elements <ID> or <URN> of the entity <AgencyOrganizationReference> narrowing down the possibilities for the unit "agentIdentifierType" to either ID or URN.

4 The DAF DDI Profile

The DDI, at its minimum configuration, does not demand any other element than the "id" attribute of the <DDIInstance> and the entity <UserID>, other elements become mandatory in a specific DDI profile as consequence of the selection of other elements they are hierarchically bound with.

Thus, the elements selected for the DAF DDI profile depended both on the characteristics this study decided in advance to attribute to the researched metadata set and on the very DDI structure.

4.1 Selecting the DDI Elements for the DAF DDI Profile

Since the first objective of this research was designing a metadata set allowing the management of a complete general curatorial lifecycle model, the fist elements which have been included in the researched DDI profile have been the mandatory PREMIS units previously mapped to DDI, that is the complex elements <Item>, <LifecycleEvent>, <AgencyOrganizationReference> and <Citation>. In particular, since the system of URN references appeared too ambitious for the expected users of the DAF tools, the element <ID> has been chosen as unique descriptor for <AgencyOrganizationReference>. Further, the inclusion of <LifecycleEvent> and <Item> in the DAF DDI profile implied that, on the one hand, their containers, the complex elements <Group> and <Archive>, were added to the selection and, on the other, that the elements <LifecycleInformation>, <Description>, <Date>, <ArchiveOrganizationReference> and <OrganizationScheme> would have been inserted to complement the description of the one, while, the elements <ArchiveSpecific>, <ArchiveOrganizationReference> and <OrganizationScheme> to refine the description of the other.

Then, aiming at improving the compatibility of the first DAF interactive tool with the data from future audits, since 11 elements out of 16 of its set are borrowed from the DCMI but the element <Purpose>, the entities of the complex <DCElements> have been added to the DAF prototype's profile as complementary descriptors for the entity <Citation> and the element <Purpose> has been inserted among the descriptors of <DDIInstance>.

But, where the DDI schemas did not influence the selection of the descriptors, some conceptual refinement were applied to specify the use of the selected elements in the researched DDI profile.

That was the case of <OrganizationScheme>, which would be validated by a parser even if not containing further descriptors but its "id" attribute; thus, the elements <OrganizationSchemeReference> and <Description>, bearing the ideal minimum information describing the organisation acting as archive, have been added to the selection.

Similarly, the descriptors contained in <DCElements>, as it is stressed in the DDI technical specifications, are not sufficient to obliterate identical DDI elements from <Citation> since the Dublin Core is not used as main citation mechanism in DDI and its use is aimed just at supporting applications which do not understand DDI; for this reason, within <Citation> have been included also <Creator> and <Copyright> while within <Item> the element <Format>. To ease the research of text documents it has been included also the element <Abstract> of <Citation>.

Other elements have been included being pertinent to the curatorial procedures implemented within research institutions and to their needs for more adequate curation and preservation policies. In effect, the elements <LocationInArchive> and <ArchiveModuleName> have been selected to complement the records concerning the ingestion process when this takes place on movable media physically stored in the repositories, as it happens in many institutions [6] In the same way, the elements <AccessTypeName>, <AccessPermission>, <Restrictions> and <AccessRestrictionDate> have been selected to offer a support for the implementation of formal procedures, while the element <EventType> of <LifecycleEvent> has been inserted to support the use of vocabularies corresponding to formal curatorial procedures. The elements <CollectionCompleteness>

and <OriginalArchiveOrganizationReference> of <Collection>, and <FundingInformation> have been selected to encourage the implementation of selection and disposal procedures through the assessment of the digital assets value.

4.2 The DAF DDI Profile

Given these premises, below is transcribed the XML resulting from the researched DDI profile:

```
<?xml version="1.0" encoding="UTF-8"?>
<DDIInstance xmlns="ddi:instance:3_1"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="ddi:conceptualcomponent:3_1
   file:/your_location_for_the_DDI_schema/instance.xsd"
 id="">
<UserID xmlns="ddi:reusable:3_1" type=""></UserID>
<Citation xmlns="ddi:reusable:3_1">
<Title></Title>
<Creator></Creator>
<Copyright></Copyright>
<DCElements xmlns="ddi:dcelements:3_1">
<title xmlns="http://purl.org/dc/elements/1.1/"></title>
<creator
xmlns="http://purl.org/dc/elements/1.1/"></creator>
<type xmlns="http://purl.org/dc/elements/1.1/"></type>
<format
xmlns="http://purl.org/dc/elements/1.1/"></format>
<date xmlns="http://purl.org/dc/elements/1.1/"></date>
<source
xmlns="http://purl.org/dc/elements/1.1/"></source>
<identifier
xmlns="http://purl.org/dc/elements/1.1/"></identifier>
<rights
xmlns="http://purl.org/dc/elements/1.1/"></rights>
<relation
xmlns="http://purl.org/dc/elements/1.1/"></relation>
<subject
xmlns="http://purl.org/dc/elements/1.1/"></subject>
<description
xmlns="http://purl.org/dc/elements/1.1/"></description>
</DCElements>
</Citation>
<Group xmlns="ddi:group:3_1" id="">
<Abstract id="">
<Content xmlns="ddi:reusable:3_1"></Content>
</Abstract>
<Purpose id="">
<Content xmlns="ddi:reusable:3_1"></Content>
</Purpose>
<Archive xmlns="ddi:archive:3_1" id="">
<ArchiveModuleName></ArchiveModuleName>
<ArchiveSpecific>
<ArchiveOrganizationReference>
<URN xmlns="ddi:reusable:3_1"></URN>
```

```
</ArchiveOrganizationReference>
<Item>
<LocationInArchive></LocationInArchive>
<Format></Format>
<Access id="">
<AccessTypeName></AccessTypeName>
<AccessPermission></AccessPermission>
<Restrictions></Restrictions>
<AccessRestrictionDate>
<SimpleDate xmlns="ddi:reusable:3_1">xxxx-xx-
xx</SimpleDate>
<User xmlns="ddi:reusable:3_1"></User>
</AccessRestrictionDate>
</Access>
</Item>
<Collection>
<OriginalArchiveOrganizationReference>
<URN xmlns="ddi:reusable:3_1"></URN>
</OriginalArchiveOrganizationReference>
<CollectionCompleteness></CollectionCompleteness>
</Collection>
<FundingInformation xmlns="ddi:reusable:3_1">
<AgencyOrganizationReference>
<ID></ID>
</AgencyOrganizationReference>
<Description></Description>
</FundingInformation>
</ArchiveSpecific>
<OrganizationScheme id="">
<OrganizationSchemeReference>
<URN xmlns="ddi:reusable:3_1"></URN>
</OrganizationSchemeReference>
</OrganizationScheme>
<LifecycleInformation xmlns="ddi:reusable:3_1">
<LifecycleEvent id="">
<EventType></EventType>
<Date>
<SimpleDate>xxxx-xx-xx</SimpleDate>
</Date>
<AgencyOrganizationReference>
<URN></URN>
</AgencyOrganizationReference>
<Description></Description>
</LifecycleEvent>
</LifecycleInformation>
</Archive>
</Group>
</DDIInstance>
```

5 The DAF DDI Profile for Cultural Heritage Institutions

Apart from supporting the adoption of policies and the implementation of procedures for the curation of digital assets, the DAF DDI profile can significantly benefit small

Cultural Heritage institutions, such as museums and government agencies, by contributing to both the enhancement of the repositories' information quality and the sustainability of the Digital Curation and Preservation practices.

5.1 Metadata Quality

DDI satisfies all the characteristics that are commonly attached to quality metadata in regards to their semantic and syntactic structures, to their data values, and to their context of use.

These characteristics concern a) the completeness of the information, b) its accuracy, that is its coherence, c) the provenance of the metadata set and its logical consistency, d) its synchronization with the instances and its maintainability, e) and the accessibility of the information in respect to its context of use. As explained above, the DDI modular structure allows for coherence control in the records, versioning and access management, and, despite being designed for the quantitative researches in social sciences, its application in a general context is still adequate because of the relatively exhaustive information conveyed. In effect, the completeness of a metadata set for the Cultural Heritage institutions should be assessed against the needs of the growing federation of digital resources in this field [7].

It has been reported that immediate local needs, such as for example disciplinary ones, have often taken priority over the needs of interoperability in Cultural Heritage institutions [8]. As a result, a tension between efficient cataloguing and quality cataloguing has become particularly evident in this context undermining the integration of resources, for example, harvested adopting the OAI protocol. The difficulties suffered by the aggregator and the service providers maintaining metadata consistency in a federated environment demonstrate that shareable metadata are crucial for the data accessibility and very detailed records are not the results of quality metadata.

The discipline unspecific approach to curation of the DAF DDI profile, the integration of DCMI elements within its descriptive metadata, and the lightness of the records produced facilitate the interoperability of this set. In adjunct, the semantic structure of the DDI DAF profile is fostered by its predisposition to implement controlled vocabularies while its syntactic consistency by the very structure of the DDI set which avoids ambiguities and incoherence in the data values.

Furthermore, the versioning control granted by DDI is particularly valuable for the institutions aiming at employing their resources in digital academic practices, such as for example the emerging non-linear scholarly publications and the e-learning initiatives.

5.2 Contribution to the Sustainability of Curatorial Practices

The digital Cultural Heritage initiatives tend to trigger preoccupations about their sustainability in small institutions; as a result the main focus of information and communication technology development in the Heritage sector concentrates on medium to larger institutions [9] The nature of this problem is mainly organisational and depends on a frequent lack of business plans and other planning tools that exposes small institutions to failure rendering operational digital cultural projects because of the resulting inadequate resources and untenable staff workloads [10].

The DAF DDI profile promotes and eases curatorial planning for the entire lifecycle of digital assets and in doing so helps improving also the internal business processes. In effect, adopting this DDI profile, small institutions would be led to create an accurate curatorial planning involving in particular the assessment of the economical value of the digital assets stored in their repositories, especially through the information recorded by <FundingInformation>, and the evaluation of the benefits obtainable from their exploitation, through the management of the intellectual rights. In this way, and taking advantage of the DDI control on data versioning, the small institutions could also implement formal practices of disposal of their obsolete and not profitable assets.

Further, the integrability of DDI in the major DMS improve the feasibility of the implementation of curatorial procedures in small institution since its users in this way can have recourse to robust products, easy to use, well serviced and stable, that is cheap to be implemented and maintained.

6 Conclusions

Reconstructing its design process, the DAF DDI profile has been demonstrated being a quality metadata for the integrated management of digital assets, that is the management along their entire lifecycle, and especially fit for the framework of federated digital resources.

A part from serving the redevelopment of the DAF interactive tool, the DAF DDI profile has been shown facilitating, in the context of the small Cultural Heritage institutions, the adoption of preservation and curation policies, of business plans, and the implementation of curatorial procedures, also through its integration in the major DMS.

Further, the DAF DDI profile is the first example of a DDI 3.1 implementation in a general curatorial context, that is not disciplinary specific, and its integration with PREMIS 2.0 is the result of the first published mapping of PREMIS to DDI.

Acknowledgments. The author would like to thank all the colleagues from HATII which offered guidance, suggestions and support for this research, and in particular Dr Ian Anderson, Patrick McCann, Joy Davidson, and Sarah Jones.

References

1. Thomas, W., Gregory, A., Gager, J., Kuo, I.-.L., Wackerow, A., Nelson, C.: Data Documentation Initiative (DDI), Technical Specification, Part 1: Overview, Version 3.1, DDI Alliance (2009)
2. Higgins, S.: Draft DCC Curation Lifecycle Model. IJDC 2(2), 82–87 (2007)
3. Martinez, L.: The Data Documentation Initiative (DDI) and Institutional Repositories (2008), http://www.disc-uk.org/docs/DDI_and_IRs.pdf
4. DDI: Data Documentation Initiative, Technical Specification, User Guide version 3.0, DDI Alliance (2008)

5. Thomas, W.: [DDI-SRG] DDI 3.0 Questions - preservation, fwd (2008), http://www.icpsr.umich.edu/pipermail/ddi-srg/2008-January/002695.html
6. Jones, S., Ross, S.: DAFD Final Report (2009), http://www.data-audit.eu/docs/DAFDfinalreport.pdf
7. Shreeves, S., Riley, J., Milewicz, L.: Moving towards shareable metadata. First Monday, 8(11) (2006), http://www.firstmonday.org/issues/issue11_8/shreeves/index.html
8. Shreeves, S., Knutson, E., Stvilia, B., Palmer, C., Twidale, M., Cole, T.: Is 'quality' metadata 'shareable' metadata? The implications of local metadata practice on federated collections. In: Proceedings of the Twelfth National Conference of the Association of College and Research Libraries, pp. 223–237. Association of College and Research Libraries, Minneapolis (2005)
9. Geser, G.: Assessing the readiness of small heritage institutions for e-culture technologies. Digi. CULT. info. 9, 8–13 (2004)
10. Zorich, D.M.: A Survey of Digital Cultural Heritage Initiatives and Their Sustainability Concerns. Council on Library and Information Resources, Washington D.C. (2003)