Package (+ Subscription Offered)

Document details

Version	Date	Key Changes	Contributors
0.1		n/a	OS
0.5		• Added filename guidance • Re-worked to clarify Subscription Offered data elements are only relevant to KB+ • Changed URI examples to use IDs rather than KB+ internal identifiers • Renamed many of the data elements for clarity • Moved to IdentifierNamespace+Identifier approach for all title identifiers, and removed ISSN,eISSN,DOI and Proprietary ID data elements • Added data element definitions for TIPPs • Moved to using a '.' as separator in PackageID	OS II
0.51		• Changed from ISO 8601 basic format to extended format for dates	OS II
0.6		• Added CSV Binding proposal 1 • Added CSV Binding proposal 2	OS II
0.61		• Minor corrections	OS II
0.7		• Added file validation guidelines	OS II DK
0.8	30-OCT-2012	• Added clearer guidance on repeatable elements	OS II DK

Table of Contents

Filenames

The filename will not be treated as semantically meaningful, and therefor there are no restrictions on the filename, and local conventions can be adopted as desired.

Purely as an example, a file naming convention might be:

ProviderID_PackageID_Date.extension

See below for definition of ProviderID and PackageID. Date would be expressed in ISO8601:2004 *extended* format, to the level of a specific day. To repeat, this is just an example naming convention, and will not have any functional role in the GOKb/KBPlus data import routines.

Encoding

Files must be UTF8 encoded

Structure

The data file is made up of two parts. Firstly a **header section** that describes the Data, the Package (and the Subscription Offered for KB+ only), using names and identifiers to refer to other GOKb/KB+ entities where necessary. Secondly a **TIPP section** that contains a list of all the TIPPs in the Package.

Compulsory elements are noted. Those elements related to Subscriptions Offered are only relevant to KBPlus. When these elements are omitted the data simply describes a package, with related TIPPs and TIs.

Header section

Contains the following elements:

Element label	Element Description	Compulsory?
DataSpecification	Type of data	Yes
SpecVersion	Specification Version	Yes
P_OrgName	Content Provider Organisation Name	Yes
P_OrgID	Content Provider Organisation ID	
P_OrgURI	Content Provider Organisation URI	
PackageName	Package Name	Yes
PackageID	Package ID	
PackageURI	Package URI	
SOName	KBPlus only: Subscription Offered Name	

Element label	Element Description	Compulsory?
SOID	KBPlus only: Subscription Offered ID	
SubURI	KBPlus only: Subscription URI	
SubTermStartDate	KBPlus only: Subscription Term Start Date	
SubTermEndDate	KBPlus only: Subscription Term End Date	
C_OrgName	KBPlus only: Consortium Organisation Name	
C_OrgID	KBPlus only: Consortium Organisation ID	
C_OrgURI	KBPlus only: Consortium Organisation URI	

TIPP section
A single package can contain multiple TIPPs, and this whole set of fields can be repeated to represent multiple TIPPs. Each TIPP can have multiple identifiers. Note also that each TIPP can relate to multiple platforms, although only a single ‘host‘ platform is allowed.

The TIPP section contains the following elements:

Element label	Element Description	Compulsory?	Repeatable?
PublicationTitle	Publication Title	Yes	
TitleIdentifierNamespace	Namespace for a title identifier		These fields are repeatable together to allow for multiple identifiers. When used, both elements must be present
TitleIdentifier	Identifier for the title belonging to the namespace in TitleIdentifierNamespace		
DateFirstPackageIssue	Date first issue online within this package		
VolumeFirstPackageIssue	Number of first volume online within this package		
NumberFirstPackageIssue	Number of first issue online within this package		
DateLastPackageIssue	Date last issue online within this package		
VolumeLastPackageIssue	Number of last volume online within this package		

Element label	Element Description	Compulsory?	Repeatable?
NumberLastPackageIssue	Number of first issue online within this package		
KBARTEmbargo	Information on access restrictions related to ‘moving wall’ access, expressed in format specified by KBART		
CoverageDepth	Coverage Depth		
CoverageNotes	Coverage Notes		
pub_OrgName	Publisher Organisation Name		
pub_OrgID	Publisher Organisation ID		
pub_OrgURI	Publisher Organisation URI		
PlatformName	Platform Name		Only a single ‘host’ platform can be defined, but multiple platforms of other types (e.g. ‘admin’) can be defined by repeating these elements
PlatformID	Platform ID		
PlatformURI	Platform URI		
PlatformRole	Platform Role		
PlatformTitleURL	Platform Title URL		

Data Element definitions

DataSpecification

A lowercase alphanumeric string which defines the type of data to be described. These are predefined by GOKb/KB+. This document describes the “package” data type.

Example: package

SpecVersion

A real number which defines the specification version of this data type. These are predefined by GOKb/KB+. This document describes version 3.0 of the package data type.

Example: 3.0

P_OrgName

The Provider Organisation Name (P_OrgName) is the name by which GOKb/KB+ refers to the Content Provider for the package.

Example: Springer

P_OrgID
The Provider Organisation ID (P_OrgID) is created by taking the P_OrgName and carrying out the following operations:
Convert to lowercase
Replace all whitespace elements with '_' (U+005F)

Example: springer

P_OrgURI
A Provider Organisation URI (P_OrgURI) should be used where it is known. For example, KB+ assigns URIs to entities.

Example: uri://kbplus/org/springer

PackageName
A Package name (PackageName) is taken directly from the provider of the package.

Example: Open Access Hybrid

PackageID
A Package ID (PackageID) is created by taking the PackageName and carrying out the following operations:
Pre-pend the Provider Name followed by "." (U+002E)
Convert to lowercase
Replace all whitespace elements with '_' (U+005F)

Example: springer.open_access_hybrid

PackageURI
A Package URI (PackageURI) should be used where it is known For example, KB+ assigns URIs to entities.

Example: uri://kbplus/package/springer.open_access_hybrid

SOName - KB+ only
A Subscription Offered name (SOName) the name by which JISC Collections refers to the Subscription, and is made of the following parts:
"[Package Name] [Subscription Term Start date]-[Subscription Term End date] [Consortium Name]"

Example: Springer Open Access 2011-2012 NESLI2

SOID - KB+ only
A Subscription Offered ID (SOID) is created by taking the SOName and carrying out the following operations:
Convert to lowercase
Replace all whitespace elements with '_' (U+005F)

Example: springer_open_access_2011-2012_nesli2

SubURI - KB+ only
A Subscription URI (SubURI) should be used where it is known For example, KB+ assigns URIs to entities.

Example: uri://kbplus/sub/springer_open_access_2011-2012_nesli2

SubTermStartDate - KB+ only
The date on which the Subscription Term starts. Must be expressed in ISO8601:2004 *extended* format, to the level of a specific day.

Example: 2011-01-01

SubTermEndDate - KB+ only
The date on which the Subscription Term starts. Must be expressed in ISO8601:2004 *extended* format, to the level of a specific day.

Example: 2012-12-31

C_OrgName - KB+ only
The Consortium Organisation Name (C_OrgName) is the name by which JISC Collections refers to the Consortium.

Example: NESLi2

C_OrgID - KB+ only
The Consortium Organisation ID (C_OrgID) is created by taking the C_OrgName and carrying out the following operations:
Convert to lowercase
Replace all whitespace elements with '_' (U+005F)

Example: nesli2

C_OrgURI - KB+ only
The Consortium Organisation URI (C_OrgURI) should be used where it is known For example, KB+ assigns URIs to entities.

Example: uri://kbplus/org/nesli2

PublicationTitle
The title of the resource described by the TIPP, equivalent to KBART publication_title, and should follow the guidelines for transcribing the title given by KBART.

Example: 3 Biotech

TitleIdentifierNamespace
A namespace (type) of an identifier to be specified in the TitleIdentifier data element. The namespace should be a lowercase alphanumeric string. Common namespaces are "issn","eissn","doi","isbn". Where a publisher, content provider or other organisation has allocated a proprietary identifier for the resource, the namespace should be created by taking the name of the issuing organisation and carrying out the following operations:
Remove all non-alphanumeric characters
Convert to lowercase

Example: issn
Example of proprietary namespace: wiley

TitleIdentifier
An identifier for the resource described by the TIPP, in the namespace specified in the TitleIdentifierNamespace data element. Where the identifier is an ISSN, KBART rules should be followed (i.e. 9 characters including hyphen and check digit).

Example: 2190-5738

DateFirstPackageIssue
The date of the first issue included in the TIPP. Must be expressed in ISO8601:2004 *extended* format, to the level of a specific day. Where no specific day is available, the first of the given month should be used. Where no specific month is available, January should be used.

Example: 2011-01-01

VolumeFirstPackageIssue
A number or label for the volume of the first issue included in the TIPP. KBART guidance for num_first_vol_online applies, which means only a number or label for the volume should be included in this field, and any terms preceding this label such as 'vol.' or 'Volume' should be ignored.

Example: 1

NumberFirstPackageIssue
A number or label for the first issue included in the TIPP. KBART guidance for num_first_issue_online applies, which means only a number or label for the volume should be included in this field, and any terms preceding this label such as 'no.' or 'Issue' should be ignored.

Example: 1

DateLastPackageIssue
The date of the last issue included in the TIPP. If specified, must be expressed in ISO8601:2004 *extended* format, to the level of a specific day. Where no specific day is available, the last day of the given month should be used. Where no specific month is available, December should be used.

The KBART convention of assuming that the lack of a DateLastPackageIssue, VolumeLastPackageIssue and NumberLastPackageIssue indicated 'up to the current issue' will be adopted.

Example: 2012-12-31

VolumeLastPackageIssue
A number or label for the volume of the last issue included in the TIPP. KBART guidance for num_last_vol_online applies, which means only a number or label for the volume should be included in this field, and any terms preceding this label such as 'vol.' or 'Volume' should be ignored.

The KBART convention of assuming that the lack of a DateLastPackageIssue, VolumeLastPackageIssue and NumberLastPackageIssue indicated ‘up to the current issue’ will be adopted.

Example: 2

NumberLastPackageIssue
A number or label for the last issue included in the TIPP. KBART guidance for num_last_issue_online applies, which means only a number or label for the volume should be included in this field, and any terms preceding this label such as ‘no.’ or ‘Issue’ should be ignored.

The KBART convention of assuming that the lack of a DateLastPackageIssue, VolumeLastPackageIssue and NumberLastPackageIssue indicated ‘up to the current issue’ will be adopted.

Example: 12

KBARTEmbargo
The KBART ‘embargo_info’ data field. This has it’s own specific syntax specified in the KBART Guidelines http://www.uksg.org/kbart/s5/guidelines/data_fields#embargo

Example: P1Y

KBARTCoverageDepth
The KBART ‘coverage_depth’ data field. This has it’s own specific syntax specified in the KBART Guidelines http://www.uksg.org/kbart/s5/guidelines/data_fields#coverage_depth

Example: fulltext

KBARTCoverageNotes
The KBART ‘coverage_notes’ data field. This is a free-text field, the details for which are specified in the KBART Guidelines http://www.uksg.org/kbart/s5/guidelines/data_fields#coverage_notes

Example: Excludes letters and book reviews

pub_OrgName
The name by which GOKb/KB+ refers to the publishing organisation

Example: Springer Berlin Heidelberg

pub_OrgID
The Publishing Organisation ID (P_OrgID) is created by taking the P_OrgName and carrying out the following operations:
Convert to lowercase
Replace all whitespace elements with ‘_’ (U+005F)

Example: springer_berlin_heidelberg

pub_OrgURI
The Publishing Organisation URI (P_OrgURI) should be used where it is known. For example, KB+ assigns URIs to entities.

Example: uri://kbplus/org/springer_berlin_heidelberg

PlatformName
The name by which GOKb/KB+ refers to the platform. A platform is defined as:

> An interface, accessible via a specific domain, that administers or delivers the content, or provides a route to the content, to the user[1]. A platform can play the role of either a 'host' or 'administrative' for a TIPP. Platforms under this definition are restricted to a single domain. Where a new domain is used, it should be regarded as a new platform.

Example: SpringerLink

PlatformID
The PlatformID is created by taking the PlatformName and carrying out the following operations:
Convert to lowercase
Replace all whitespace elements with '_' (U+005F)

Example: springerlink

PlatformURI
The PlatformURI should be used where it is known. For example, KB+ assigns URIs to entities. N.B. This is not the same as the URL for the platform.

Example: uri://kbplus/platform/springerlink

PlatformRole
The role (host or administrative) played by the platform in relation to the TIPP.

Example: host

PlatformTitleURL
The url of the TIPP on the specified platform.

Example: http://www.springerlink.com/content/2190-5738/

Bindings

JSON binding

It is assumed that a JSON file containing GOKb/KB+ data will be an array of objects, each object representing a data type. This would allow, for example, multiple packages to be defined in a single JSON file.

---

[1] This is based on the COUNTER Code of Practice v3 definition, but has been extended to encompass the concept of 'administraiton' separate from delivery, and to include 'gateway' as a type of platform

Each data element described above is treated as a name/value pair. Where sets of elements are repeatable (specifically TIPPs, TitleIdentifiers and Platforms) these are named arrays containing one object per set of elements.

Sample:


```
[{
	"DataSpecification": "package",
	"SpecVersion": "3.0",
	"P_OrgName": "Springer",
	"P_OrgID": "springer",
	"P_OrgURI": "uri://kbplus/org/springer",
	"PackageName": "Open Access Hybrid",
	"PackageID": "springer.open_access_hybrid",
	"SOName": "Springer Open Access 2011-2012 NESLI2",
	"SOID": "springer_open_access_2011-2012_nesli2",
	"SubTermStartDate": "2011-01-01",
	"SubTermEndDate": "2012-12-31",
	"C_OrgName": "NESLi2",
	"C_OrgID": "nesli2",
	"C_OrgURI": "uri://kbplus/org/nesli2",
	"TIPPs": [{
		"PublicationTitle": "3 Biotech",
		"TitleIdentifiers": [{
			"TitleIdentifierNamespace": "eissn",
			"TitleIdentifier": "2190-5738"
		}],
		"DateFirstPackageIssue": "2011-01-01",
		"VolumeFirstPackageIssue": "1",
		"pub_OrgName": "Springer Berlin Heidelberg",
		"pub_OrgID": "springer_berlin_heidelberg",
		"pub_OrgURI": "uri://kbplus/org/springer_berlin_heidelberg",
		"Platforms": [{
			"PlatformName": "SpringerLink",
			"PlatformID": "springerlink",
			"PlatformURI": "uri://kbplus/platform/springerlink",
			"PlatformRole": "host",
			"platform_title_url": "http://www.springerlink.com/content/2190-5738/"
		}, {
			"PlatformName": "SpringerLink",
			"PlatformID": "springerlink",
			"PlatformURI": "uri://kbplus/platform/springerlink",
			"PlatformRole": "administrative"
		}]
	}, {
		"PublicationTitle": "Advances in Difference Equations",
		"TitleIdentifiers": [{
			"TitleIdentifierNamespace": "eissn",
			"TitleIdentifier": "1687-1847"
		}],
		"DateFirstPackageIssue": "2011-01-01",
		"VolumeFirstPackageIssue": "2011",
		"pub_OrgName": "Springer International Publishing, CH",
```

```
                "pub_OrgID": "springer_international_publishering,_ch",
                "pub_OrgURI": "uri://kbplus/org/springer_international_publishering,_ch",
                "Platforms": [{
                        "PlatformName": "SpringerLink",
                        "PlatformID": "springerlink",
                        "PlatformURI": "uri://kbplus/platform/springerlink",
                        "PlatformRole": "host",
                        "platform_title_url": "http://www.springerlink.com/content/1687-1847/"
                }, {
                        "PlatformName": "SpringerLink",
                        "PlatformID": "springerlink",
                        "PlatformURI": "uri://kbplus/platform/springerlink",
                        "PlatformRole": "administrative"
                }]
        }]
}]
```


JSON Validation
Files should be validated as follows:

Check for valid UTF8 encoding
Check for valid JSON

This can be done by any relevant tools. As an example Ubuntu includes a jsonlint command that can check both encoding and format, see http://manpages.ubuntu.com/manpages/lucid/man1/jsonlint.1.html for details.

CSV binding - Proposal 1

row	content
1	Column names for **header section**. The names used should be the data element names
2	Data header information
3	Column names for **TIPP section**. The names used should be the data element names
4 onwards	Rows representing TIPPs. There can be multiple lines per TIPP to accommodate multiple IDs and multiple platforms. Lines are identified as belonging to the same TIPP by an additional TIPPNumber field

Values containing commas should be enclosed in inverted commas. Where values do not contain commas use of inverted comma is optional.

Sample:

DataSpecification,SpecVersion,P_OrgName,P_OrgID,P_OrgURI,PackageName,PackageID,PackageURI,SOName,SOID,
SubURI,SubTermStartDate,SubTermEndDate,C_OrgName,C_OrgID,C_OrgURI,,,,,

Package,3,Springer,springer,uri://kbplus/org/springer,Open Access Hybrid,springer.open_access_hybrid,,Springer Open Access 2011-2012 NESLI2,springer_open_access_2011-2012_nesli2,,2011-01-01,2012-12-31,NESLi2,nesli2,uri://kbplus/org/nesli2,,,,,
TIPPNumber,PublicationTitle,TitleIdentifierNamespace,TitleIdentifier,DateFirstPackageIssue,VolumeFirstPackageIssue,NumberFirstPackageIssue,DateLastPackageIssue,VolumeLastPackageIssue,NumberLastPackageIssue,KBARTEmbargo,CoverageDepth,CoverageNotes,pub_OrgName,pub_OrgID,pub_OrgURI,PlatformName,PlatformID,PlatformURI,PlatformRole,PlatformTitleURL
1,3 Biotech,eissn,2190-5738,2011-01-01,1,,,,,,,,Springer Berlin Heidelberg,springer_berlin_heidelberg,uri://kbplus/org/springer_berlin_heidelberg,SpringerLink,springerlink,uri://kbplus/platform/springerlink,host,http://www.springerlink.com/content/2190-5738/
1,,springer,13205,,,,,,,,,,,,SpringerLink,springerlink,uri://kbplus/platform/springerlink,administrative,
2,Advances in Difference Equations,eissn,1687-1847,2011-01-01,2011,,,,,,,,"Springer International Publishing, CH","springer_international_publishering,_ch","uri://kbplus/org/springer_international_publishering,_ch",SpringerLink,springerlink,uri://kbplus/platform/springerlink,host,http://www.springerlink.com/content/1687-1847/
2,,,,,,,,,,,,,,,,SpringerLink,springerlink,uri://kbplus/platform/springerlink,administrative,

CSV binding - Proposal 2

row	content
1	Column names for **header section**. The names used should be the data element names
2	Data header information
3	Column names for **TIPP section.** The names used should be the data element names, with the following exceptions: • TitleIdentifierNamespace • TitleIdentifier Instead of these two columns, there is a single column per identifier with the following naming convention: TitleIdentifier.[TitleIdentifierNamespace] Example: TitleIdentifier.eissn Example: TitleIdentifier.doi • PlatformName • PlatformID • PlatformURI • PlatformRole • PlatformTitleURL Instead of these five columns, there is a single column per platform the header ‘platform’. The five values should be stored in this field as comma separated values as detailed below.

row	content
4 onwards	Rows representing TIPPs. There is a single line per TIPP. Multiple identifiers are represented as multiple columns using the naming convention stated above. Multiple platforms are represented as multiple columns, one per platform. The platform column should contain all the values: • PlatformName • PlatformID • PlatformURI • PlatformRole • PlatformTitleURL As comma separated values.

Values containing commas should be enclosed in inverted commas. Where values do not contain commas use of inverted comma is optional.

Sample:

DataSpecification,SpecVersion,P_OrgName,P_OrgID,P_OrgURI,PackageName,PackageID,PackageURI,SOName,SOID,SubURI,SubTermStartDate,SubTermEndDate,C_OrgName,C_OrgID,C_OrgURI,,,,,
Package,3,Springer,springer,uri://kbplus/org/springer,Open Access Hybrid,springer.open_access_hybrid,,Springer Open Access 2011-2012 NESLI2,springer_open_access_2011-2012_nesli2,,2011-01-01,2012-12-31,NESLi2,nesli2,uri://kbplus/org/nesli2,,,,,
PublicationTitle,TitleIdentifier.eissn,TitleIdentifier.springer,DateFirstPackageIssue,VolumeFirstPackageIssue,NumberFirstPackageIssue,DateLastPackageIssue,VolumeLastPackageIssue,NumberLastPackageIssue,KBARTEmbargo,CoverageDepth,CoverageNotes,pub_OrgName,pub_OrgID,pub_OrgURI,Platform,Platform
3 Biotech,2190-5738,13205,2011-01-01,1,,,,,,,,Springer Berlin Heidelberg,springer_berlin_heidelberg,uri://kbplus/org/springer_berlin_heidelberg,"SpringerLink,springerlink,uri://kbplus/platform/springerlink,host,http://www.springerlink.com/content/2190-5738/","SpringerLink,springerlink,uri://kbplus/platform/springerlink,administrative,"
Advances in Difference Equations,eissn,1687-1847,2011-01-01,2011,,,,,,,,"Springer International Publishing, CH","springer_international_publishering,_ch","uri://kbplus/org/springer_international_publishering,_ch","SpringerLink,springerlink,uri://kbplus/platform/springerlink,host,http://www.springerlink.com/content/1687-1847/","SpringerLink,springerlink,uri://kbplus/platform/springerlink,administrative,

CSV Validation
While it is not possible to validate the csv file using generic tools, files should be checked for valid UTF8 encoding. This can be done by any relevant tools. As an example the GNU utility ‘iconv’ could be used.