Title	CLARIN Federated Content Search (CLARIN-FCS) – Data Views
Version	1.0
Authors	Oliver Schonefeld (IDS, CLARIN-D) (Editor) Thomas Eckart (ASV Leipzig, CLARIN-D) Thomas Kisler (BAS, CLARIN-D) Christoph Draxler (BAS, CLARIN-D) Kai Zimmer (BBAW, CLARIN-D,) Matej Ďurčo (CLARIN Center Vienna, CLARIN-AT) Yana Panchenko (Tübingen University, CLARIN-D) Hanna Hedeland (HZSK, CLARIN-D) Andre Blessing (Stuttgart University, CLARIN-D) Olha Shkaravska (MPI for Psycholinguistics, CLARIN-NL)
Date	2014-04-07
Status	Draft for approval by SCCTC
Distribution	SCCTC
ID	CE-2014-0317

1 Introduction

This specification is a supplementary specification to the CLARIN-FCS Core specification and defines additional Data View to be used in CLARIN-FCS. This specification defines the supplementary Data Views. For detailed information about the CLARIN-FCS *interface specification*, see [CLARIN-FCS-Core].

1.1 Terminology

The key words `MUST`, `MUST NOT`, `REQUIRED`, `SHALL`, `SHALL NOT`, `SHOULD`, `SHOULD NOT`, `RECOMMENDED`, `MAY`, and `OPTIONAL` in this document are to be interpreted as described in [RFC2119].

1.2 Normative References

RFC2119

> Key words for use in RFCs to Indicate Requirement Levels, IETF RFC 2119, March 1997, http://www.ietf.org/rfc/rfc2119.txt

XML-Namespaces

> Namespaces in XML 1.0 (Third Edition), W3C, 8 December 2009, http://www.w3.org/TR/2009/REC-xml-names-20091208/

CLARIN-FCS-Core

> CLARIN Federated Content Search (CLARIN-FCS) - Core, SCCTC FCS Task-Force, March 2014, http://www.clarin.eu/fcs/add/link/here

1.3 Non-Normative References

RFC6838

Media Type Specifications and Registration Procedures, IETF RFC 6838, January 2013, http://www.ietf.org/rfc/rfc6838.txt

RFC3023

XML Media Types, IETF RFC 3023, January 2001, http://www.ietf.org/rfc/rfc3023.txt

KML

Keyhole Markup Language (KML), Open Geospatial Consortium, 2008, http://www.opengeospatial.org/standards/kml

1.4 Typographic and XML Namespace conventions

The following typographic conventions for XML fragments will be used throughout this specification:

- `<prefix:Element>`
  An XML element with the Generic Identifier *Element* that is bound to an XML namespace denoted by the prefix *prefix*.
- `@attr`
  An XML attribute with the name *attr*
- `string`
  The literal *string* must be used either as element content or attribute value.

Endpoints and Clients MUST adhere to the [XML-Namespaces] specification. The CLARIN-FCS interface specification generally does not dictate whether XML elements should be serialized in their prefixed or non-prefixed syntax, but Endpoints MUST ensure that the correct XML namespace is used for elements and that XML namespaces are declared correctly. Clients MUST be agnostic regarding syntax for serializing the XML elements, i.e. if the prefixed or un-prefixed variant was used, and SHOULD operate solely on *expanded names*, i.e. pairs of *namespace name* and *local name*.

The following XML namespace names and prefixes are used throughout this specification. The column "Recommended Syntax" indicates which syntax variant SHOULD be used by the Endpoint to serialize the XML response.

Prefix	Namespace Name	Comment	Recommended Syntax
fsc	http://clarin.eu/fcs/resource	CLARIN-FCS Resources	prefixed
cmdi	http://www.clarin.eu/cmd/	Component Metadata	un-prefixed
kml	http://www.opengis.net/kml/2.2	Keyhole Markup Language	un-prefixed

2 Data Views

A Data View serves as a container for representing search results within CLARIN-FCS. Data Views are designed to allow for different representations of results. This specification defines supplementary

Data Views beyond the Generic Hits Data View, which is defined as part of the CLARIN-FCS Core specification. For detailed information as to what Data Views are and how they are integrated in CLARIN-FCS, see [CLARIN-FCS-Core].

NOTE: The examples in the following sections *show only* the payload with the enclosing `<fcs:DataView>` element of a Data View. Of course, the Data View must be embedded either in a `<fcs:Resource>` or a `<fcs:ResourceFragment>` element. The `@pid` and `@ref` attributes have been omitted for all *inline* payload types.

2.1 Generic Hits (HITS)

The *Generic Hits* (HITS) Data View is an integral part of the Core specification and serves as the as the *most basic* agreement in CLARIN-FCS for the serialization of search results. For details about this Data View, see refer to the Core specification [CLARIN-FCS-Core, Section "Generic Hits (HITS)"].

2.2 Component Metadata (CMDI)

Description	A CMDI metadata record
MIME type	`application/x-cmdi+xml`
Payload Disposition	*inline* or *reference*
Payload Delivery	*send-by-default* (`RECOMMENDED`)
Recommended Short Identifier	`cmdi` (`RECOMMENDED`)

The *Component Metadata* Data View allows the embedding of a CMDI metadata record that is *applicable* to the specific context into the Endpoint response, e.g. metadata about the resource in which the hit was produced. If this CMDI record is applicable for the entire Resource, it `SHOULD` be put in a `<fcs:DataView>` element below the `<fcs:Resource>` element. If it is applicable to the Resource Fragment, i.e. it contains more specialized metadata than the metadata for the encompassing resource, it `SHOULD` be put in a `<fcs:DataView>` element below the `<fcs:ResourceFragment>` element. Endpoints `SHOULD` provide the payload *inline*, but Endpoints `MAY` also use the *reference* method. If an Endpoint uses the *reference* method, the CMDI metadata record `MUST` be downloadable without any restrictions.

```
<!-- potential @pid and @ref attributes omitted -->
<fcs:DataView type="application/x-cmdi+xml">
  <cmdi:CMD xmlns:cmdi="http://www.clarin.eu/cmd/" CMDVersion="1.1">
    <!-- content omitted -->
  </cmdi:CMD>
</fcs:DataView>
```

Example 2: CMDI Data View deposited “inline”.

```
<!-- potential @pid attribute omitted -->
<fcs:DataView type="application/x-cmdi+xml"
              ref="http://repos.example.org/resources/4711/0815.cmdi" />
```

Example 1: CMDI Data View deposited by “reference”

2.3 Images (IMG)

Description	An image related to the hit
MIME type	`image/png`, `image/jpeg`, `image/gif`, `image/svg+xml`
Payload Disposition	*reference*
Payload Delivery	*need-to-request* (RECOMMENDED)
Recommended Short Identifier	`image` (RECOMMENDED)

The *Image* Data View provides an image that is relevant to the hit, e.g. a facsimile of the source of a transcription. Endpoints MUST provide the payload by the *reference* method and the image file SHOULD be downloadable without any restrictions.

```
<!-- potential @pid attribute omitted -->
<fcs:DataView type="image/png"
              ref="http://repos.example.org/resources/4711/0815.png" />
```

Example 3: Image Data View

2.4 Geolocation (GEO)

Description	An geographic location related to the hit
MIME type	`application/vnd.google-earth.kml+xml`
Payload Disposition	*inline*
Payload Delivery	*need-to-request* (RECOMMENDED)
Recommended Short Identifier	`kml` (RECOMMENDED)

The *Geolocation* Data View allows to geolocalize a hit. If MUST be encoded using the XML representation of the Keyhole Markup Language (KML). The KML fragment MUST comply with the specification as defined by [KML].

```
<!-- potential @pid and @ref attributes omitted -->
<fcs:DataView type="application/vnd.google-earth.kml+xml">
  <kml:kml xmlns:kml="http://www.opengis.net/kml/2.2">
    <kml:Placemark>
      <kml:name>IDS Mannheim</kml:name>
      <kml:description>Institut für Deutsche Sprache, R5 6-13,↩
                                    68161 Mannheim, Germany</kml:description>
      <kml:Point>
        <kml:coordinates>8.4719510,49.4883700,0</kml:coordinates>
      </kml:Point>
    </kml:Placemark>
  </kml:kml>
</fcs:
```

Example 4: Geolocation Data View