

DINI-nestor AG Forschungsdaten: PIDs

Session 2 für Fortgeschrittene: Granularität, Versionierung
und dynamische Datensätze

Ulrich Schwardmann

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
(GWDG)

Am Fassberg, 37077 Göttingen
ulrich.schwardmann [at] gwdg.de

29 March 2017, Hannover

Content

DINI-nestor
Session 2

Ulrich
Schwardmann

1 Einführung Technik

Allgemein
Auflösung (Resolution)
Management

Einführung Technik

Allgemein
Auflösung
(Resolution)
Management

2 Granularität

Templates oder Fragment Identifier

Granularität

Templates oder
Fragment
Identifier

3 Types

PID Information Types
Data Type Registries

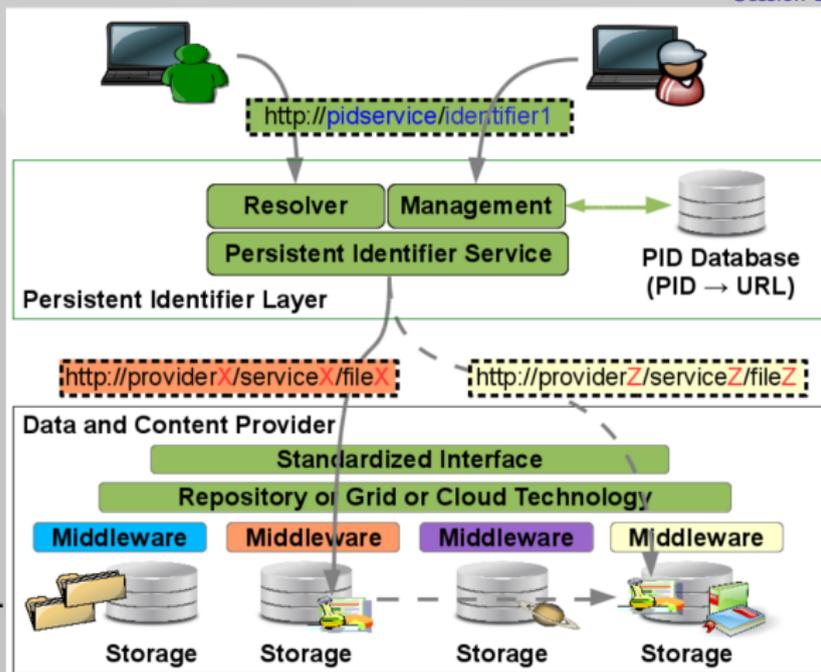
Types

PID Information
Types
Data Type
Registries

Persistent Identifier

DINI-nestor
Session 2

- Definierte Schicht, welche zwischen Referenz und Objekt vermittelt.
- Die Referenz wird beim PID Service aufgelöst, welcher zur aktuellen Position des Objekts vermittelt.
- Sollte das Objekt nicht mehr existieren, kann dies ebenfalls im PID hinterlegt werden.
- Aufbau von Mehrwertdiensten



Dokumentation und Beispiele

■ Dokumentation:

- ePIC FAQ:
http://www.pidconsortium.eu/?page_id=1060
- ePIC Policy & QoS:
http://www.pidconsortium.eu/?page_id=904
- ePIC API Dokumentation:
<http://docs.pidconsortium.eu/>

■ Beispiele für Handle PIDs

- 21.T11998/4B7335BF-CA0C-4671-9AAF-F36AE8AC88AB
- 21.T11998/TEMPLATEHANDLE
- 10.1000/182
- 11148/0000-0003-0B00-1
- 10.5072/11148/0000-0003-0B00-1

■ Demo ePIC PID Service

- <http://demo.pidconsortium.eu/handles/21.T11998/>
- User: (auf Anfrage)
- Passwd: (auf Anfrage)

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

Resolution (1)

Nutzung des Handle Proxy server, um die Beispiel PIDs oben aufzulösen

- 1 Online Form auf der web page des Handle Proxy.
- 2 Nutze Handle Proxy direkt, um PID aufzulösen (PID ist Teil der URL).
- 3 Nutze Handle Proxy, um die Metadaten der Beispiel PIDs oben zu bekommen. (Hinweis: Dokumentation des Proxy)

Handle Proxy:

- URL: <http://hdl.handle.net/>
- User/Pwd: öffentlich, also ohne Passwort

Resolution (2)

Nutze den ePIC PID Service (DEMO), um das Beispiel PID
21.T11998/0000-0011-2DA4-2 aufzulösen

- Nutze Web Browser um die Metadaten der Beispiel PIDs oben zu bekommen.
 - Hinweis: PID ist Teil der URL
- Nutze (Command Line) HTTP Client
 - Hinweis: Dokumentation der ePIC-API-v2

ePIC PID Service (DEMO)

- URL:
`http://demo.pidconsortium.eu/handles/21.T11998/`
- User: (auf Anfrage)
- Passwd: (auf Anfrage)

Lösungen (1)

- URL: `http://hdl.handle.net/21.T11998/0000-0011-2DA4-2?noredirect=1`

Handle Values for: 21.T11998/0000-0011-2DA4-2

Index	Type	Timestamp	Data
1	URL	2016-08-30 14:35:47Z	http://www.gwdg.de/TEST/099
100	HS ADMIN	2016-08-30 14:35:47Z	handle=21.PERS.0000-0011-2D9B-D; index=300; [create hdl,delete hdl,read val,modify val,del val,add

[Handle Proxy Server Documentation](#)
[Handle.net Web Site](#)

Please contact hdladmin@cnri.reston.va.us for your handle questions and comments.

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein

Auflösung
(Resolution)

Management

Granularität

Templates oder
Fragment
Identifizier

Types

PID Information
Types

Data Type
Registries

Lösungen (2)

DINI-nestor
Session 2

■ URL:

<http://pid.gwdg.de/handles/21.T11998/0000-0011-2DA4-2>

Ulrich
Schwardmann

idx	type	parsed data	data	timestamp	ttl type	ttl	refs	privs
1	URL	http://www.gwdg.de/TEST/089	ahfR0cDovL3d3cy9nd2RnLmRlL1RFU1QvMDk5	2016-08-30T14:35:47Z	0	86400		rwl-
100	HS_ADMIN	adminId 21.PERS/0000-0011-2DA4-D adminIndex 300 perms add_handle true delete_handle true add_derived_prefix false delete_derived_prefix false modify_value true remove_value true add_value true modify_admin true remove_admin true add_admin true read_value true list_handles false	B/MAAAAYMjEwLjEVSUyBwMDAwLTAwMTEBMkQ6Q1EAAABLA=	2016-08-30T14:35:47Z	0	86400		rwl-

Einführung
Technik

Allgemein

Auflösung
(Resolution)

Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types

Data Type
Registries

Management

DINI-nestor
Session 2

Ulrich
Schwardmann

Nutze REST client (z.B. curl command line client) um PIDs mit dem ePIC PID Service zu bearbeiten

- 1 Erzeuge PID mit dem Test User
 - 1 z.B. mit URL <http://www.gwdg.de/TEST/001>
- 2 Zeige die Metadaten des PID an
- 3 Ändere die PID
 - 1 z.B. Update der URL

ePIC PID Service (DEMO)

- URL:
<http://demo.pidconsortium.eu/handles/21.T11998/>
 - User: (auf Anfrage)
 - Passwd: (auf Anfrage)

Einführung
Technik

Allgemein
Auflösung
(Resolution)

Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

Lösungen (1)

```
curl -v -u 'demo1:PaSsWoRd' -H
'Accept:application/json' -H 'Content-Type:
application/json' -X POST --data ' [{"type":"URL",
"parsed_data":"http://os.helmholtz.de/de/
bewusstsein-schaerfen/workshops/rda-de-16/"} ] '
http://demo.pidconsortium.eu/handles/21.T11998
```

```
uschwarl@pcsw:~> curl -v -u 'demo1:PaSsWoRd' -H "Accept:application/json" -H "Content-Type:application/json" -X POST --data ' [{"type":"URL", "parsed_data":"http://os.helmholtz.de/de/bewusstsein-schaerfen/workshops/rda-de-16/"} ]' http://demo.pidconsortium.eu/handles/21.T11998
```

```
* Trying 134.76.30.193...
* Connected to demo.pidconsortium.eu (134.76.30.193) port 80 (#0)
* Server auth using Basic with user 'rdademo1'
> POST /handles/21.T11998 HTTP/1.1
> Host: demo.pidconsortium.eu
> Authorization: Basic cmRhZGVtbzE6VHJhaw5pbmcyMDE2
> User-Agent: curl/7.42.1
> Accept:application/json
> Content-Type:application/json
> Content-Length: 102
>
* upload completely sent off: 102 out of 102 bytes
< HTTP/1.1 201 Created
< Date: Tue, 21 Mar 2017 11:27:03 GMT
< Content-Length: 43
< Content-Type: application/json
< Location: https://demo.pidconsortium.eu/handles/21.T11998/0000-0001-3219-B
< ETag: "1B2M2Y8AsgTpgAmY7PhCfg"
< Last-Modified: Thu, 01 Jan 1970 00:00:00 GMT
<
{"epic-pid":"21.T11998/0000-0001-3219-B"}
* Connection #0 to host demo.pidconsortium.eu left intact
```

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)

Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

Lösungen (2)

```
curl -v -u 'demo1:PaSsWoRd' -H  
'Accept:application/json' -X GET
```

```
http://demo.pidconsortium.eu/handles/21.T11998/0000-0001-3219-B
```

```
uschwarl@pcscw:~> curl -v -u 'demo1:PaSsWoRd' -H 'Accept:application/json' -X GET http://  
/demo.pidconsortium.eu/handles/21.T11998/0000-0001-3219-B  
* Trying 134.76.30.193...  
* Connected to demo.pidconsortium.eu (134.76.30.193) port 80 (#0)  
* Server auth using Basic with user 'rdademo1'  
> GET /handles/21.T11998/0000-0001-3219-B HTTP/1.1  
> Host: demo.pidconsortium.eu  
> Authorization: Basic cmRhZGVtbzE6VHJhaW5pbmcyMDE2  
> User-Agent: curl/7.42.1  
> Accept:application/json  
>  
< HTTP/1.1 200 OK  
< Date: Tue, 21 Mar 2017 11:58:45 GMT  
< Content-Type: application/json  
< ETag: "VdC7i1ZImDhRQ+yjEn0agQ"  
< Last-Modified: Tue, 21 Mar 2017 11:27:03 GMT  
< Transfer-Encoding: chunked  
<  
<  
[  
{  
  "idx":1,  
  "type":"URL",  
  "parsed_data":"http://os.helmholtz.de/de/bewusstseins-schae-rfen/workshops/rda-de-16/",  
  "data":{"aHR0cDovL29zLmhmLmG1ob2x0ei5kZS9kZS9iZXdlc3N0c2Vpb1lzY2hhZXJmLnZw4vd29ya3Nob3BzL3JkYS1kZ  
S0xNi8="},  
  "timestamp":"2017-03-21T11:27:03Z",
```

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)

Management

Granularität

Templates oder
Fragment
Identifer

Types

PID Information
Types
Data Type
Registries

Granularität

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

Templates oder Fragment Identifier

Regelwerk für Zeichenketten angehängt an PIDs (urspr. für URLs, see IETF RFC 6570)

- Die Template-Implementierung im Handle-System ist eine einfache Ersetzungsregel.
- Delimiter und Ersatzzeichen sind im Prefix festgelegt.
- *Beispiel*
 - Delimiter ist @, was durch ? ersetzt wird
11858/00-ZZZZ-0000-0001-CCD1-4@aaa=bbb&ccc=ddd
 - wird übersetzt in:
`http://wwwuser.gwdg.de/~tkalman/downloads/formtest.php?aaa=bbb&ccc=ddd`
- aber Vorsicht: Fragment Identifier sind wesentlich **weniger persistent** als PIDs selbst

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

Fragment Identifiers / Template Handles

DINI-nestor
Session 2

Ulrich
Schwardmann

Hands-On Übung:

- Welcher Delimiter wird für Templates in 21.T11998 verwendet und mit welcher Zeichenfolge ersetzt.
 - (Hinweis: siehe im Inhalt von 0.NA/21.T11998 auf dem Proxy nach)
- Nutze diesen Delimiter, um ein Fragment mit dem PID 21.T11998/TEMPLATEHANDLE zu adressieren
 - (Hinweis: Löse 21.T11998/TEMPLATEHANDLEabc=123&defghi=56789@ auf dem Proxy auf

Handle Proxy:

- URL: <http://hdl.handle.net/>
- User/Pwd: öffentlich, also ohne Passwort

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifer

Types

PID Information
Types
Data Type
Registries

Lösung (4)

Index	Type	Timestamp	Data
1	HS_SITE	2016-02-05 14:21:21Z	000102030003800200000000000000100000004646573630000
2	EMAIL	2016-02-04 14:53:04Z	handle@gwdg.de
3	HS_NAMESPACE	2016-02-04 14:53:04Z	<namespace> <template delimiter="@"> <foreach> <if value="type" test="equals" expression="URL"> <value data="{data}"?\${extension}"/> </if> <else> <if value="type" test="equals" expression="HS_ALIAS"> <value data="{data}@\${extension}"/> </if> <else> <value/> <else> </else> </foreach> </template> </namespace>
100	HS_ADMIN	2016-02-04 14:53:04Z	handle=0.NA/21.ADMINLIST; index=200; [create hdl,delete hdl
101	HS_ADMIN	2016-08-30 08:45:42Z	handle=21.T11998/ADMINLIST; index=200; [create hdl,delete hdl
102	HS_ADMIN	2016-08-30 08:45:42Z	handle=21.T11998/USERLIST; index=200; [create hdl,delete hdl
200	HS_VLIST	2016-08-30 08:45:42Z	200:21.T11998/ADMINLIST
300	HS_PUBKEY	2016-02-10 14:34:19Z	0000000B4453415F5055425F4B455900000000015009760508

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

PID Information Types

DINI-nestor
Session 2

Ulrich
Schwardmann

- sind zusätzliche Metadaten in der PID-Datenbank
- direkt zugreifbar, ohne weitere Redirektion
- typische Beispiele sind
 - Checksum
 - Mime Type (inkl. Version)
 - Embargo-Zeit
 - Verfallszeiten (Löschung der Daten)
 - Versionen
 - Versionsnummer
 - Versionsrelationen: `isPreviousVersionOf`, `isNewVersionOf`
 - zusätzl. Metadaten-URLs
 - Basic Dublin Core
 - ...

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifizier

Types

PID Information
Types
Data Type
Registries

Erzeuge PID mit DC5 Metadaten

DINI-nestor
Session 2

Ulrich
Schwardmann

'Basic' PID

type	data
URL	http://www.gwdg.de/TEST/001

PID

with DC5 Metadaten

type	data
URL	http://ratswd.de/dl/downloads/langzeitarchivierung_von_forschungsdaten.pdf
TITLE	European Persistent Identifier Consortium – PIDs fuer die Wissenschaft
AUTHORS	Tibor Kálmán, Daniel Kurzawe, Ulrich Schwardmann
PUBDATE	2012
PUBLISHER	Scivero
METADATA_URL	http://d-nb.info/1030352194/about/rdf

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifizier

Types

PID Information
Types
Data Type
Registries

Type Management

Verwende REST client (e.g. curl command line client) um PIDs mit dem ePIC PID Service zu bearbeiten

- 1 Erzeuge PID mit dem Test-Nutzer mit den DC5 Metadaten
- 2 Zeige die Metadaten des PID an.

Use a REST client (e.g. curl command line client) to manage PIDs with the ePIC PID Service

- 1 Create a PID with your test user
- 2 Show the metadata of your PID

ePIC PID Service (DEMO)

- URL:
`http://demo.pidconsortium.eu/handles/21.T11998/`
- User: (wird am Ende der Session verteilt)
- Passwd: (wird am Ende der Session verteilt)

DINI-nesstor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

Lösungen (3)

```
curl -v -u 'demo1:PaSsWoRd' -H
'Accept:application/json' -H
"ContentType:application/json" -X POST --data
' [{"type":"URL",
"parsed_data":"http://ratswd.de/dl/...pdf"},
{"type":"TITLE", "parsed_data":"European
Persistent Identifier Consortium - PIDs ..."},
{"type":"AUTHORS", "parsed_data":"Tibor Kálmán,
Daniel Kurzawe, Ulrich Schwardmann"},
{"type":"PUBDATE", "parsed_data":"2012"},
{"type":"PUBLISHER", "parsed_data":"Scivero"},
{"type":"METADATA_URL", "parsed_data":
"http://d-nb.info/1030352194/about/rdf"}] '
http://demo.pidconsortium.eu/handles/21.T11998
```

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types

Data Type
Registries

Data Type Registries

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifier

Types

PID Information
Types

**Data Type
Registries**

Data Type Registries

- PID Information Types (PITs) sind bisher nicht interoperabel.
- Nun werden PIT Definitionen zunehmend in Data Type Registries (DTRs) gehalten.
- es gibt bereits eine Reihe davon
 - die auf einer Software namens *Cordra*¹ beruhen (RDA WG outcome).
 - Verwendet wird ein spezielles Vokabular für Type Spezifikationen.
 - Es gibt Planungen zur Föderation solcher DTRs.
- Föderierte DTRs können Metadaten-Interoperabilität signifikant erhöhen.

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

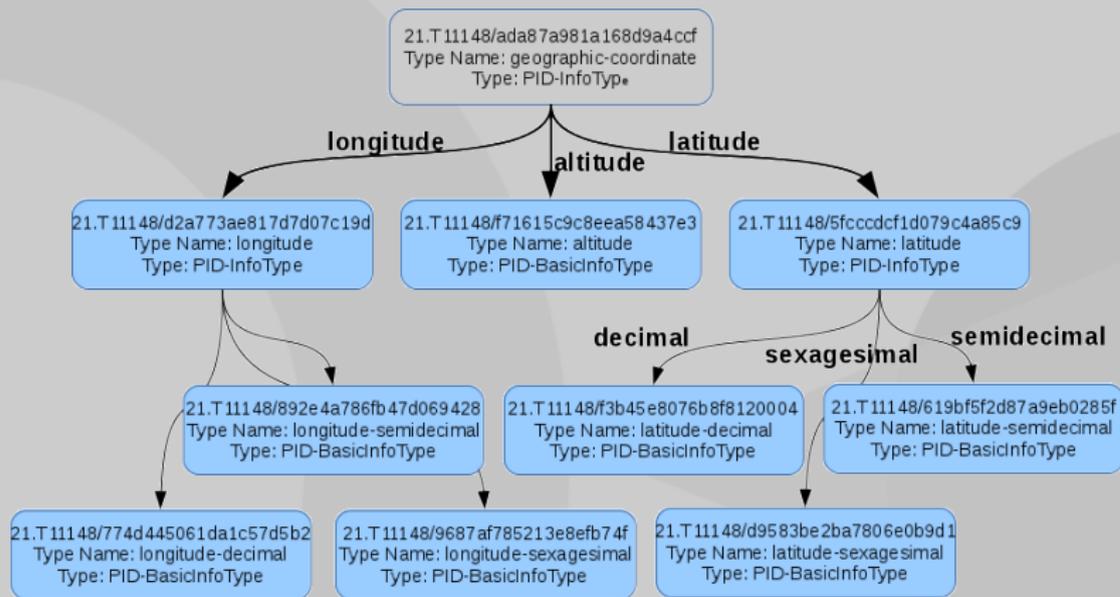
Templates oder
Fragment
Identifier

Types

PID Information
Types
Data Type
Registries

Hierarchien in Metadaten

Beispiel: geographische Koordinaten.



- Die ePIC DTR kann solche Hierarchien durch Erweiterungen des Definitions-Vokabulars ausdrücken

Die ePIC DTR Startseite

Beispiele unter: <http://dtr.pidconsortium.eu/>

- PID InfoType Zustände sind:
 - *in preparation (21.T11148), candidate, approved, deprecated (21.11104)*

Startbildschirm der PID InfoType Registry for *preparation*:

<http://dtr-test.pidconsortium.eu/>

DINI-nestor
Session 2

Ulrich
Schwardmann

Einführung
Technik
Allgemein
Auflösung
(D...)

Document Repository - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Document Repository x +

vm11.pid.gwdg.de:8081/#urls/intro.html Suchen

Document Repository Introduction All Types Sign in

Search Search

A few things to note before you get started...

Administration

You can administer the repository by accessing <http://localhost:8080/admin.html>. By default, you can administer the repository using username 'admin' and password 'changeit'. To change these defaults edit the file `repository/data/password.dct` and restart the repository.

The administration page allows you to configure metadata schemas, declare authentication and access control policies, enable federation, etc.

You can add users and groups using the main repository page (link below). Authenticate as admin and create user and group records as needed.

By default the repository is configured to store documents and associate metadata with those documents.

You have been temporarily allotted a Handle prefix 21.T11148 that will be used by the repository to allot persistent identifiers to registered digital objects.

Repository Access

You and other users can access the repository at <http://localhost:8080/> or using DOIIP at port 8900. Depending on the access controls set in the administration page, you and other users may need to authenticate prior to searching, retrieving, or registering any data or metadata.

23/26

ePIC DTR Abfragen: PID-InfoTypes

http://dtr-test.pidconsortium.eu/
#objects/?query=type:"PID-InfoType"

DINI-nestor
Session 2

The screenshot shows a Mozilla Firefox browser window displaying the Document Repository website. The address bar contains the URL: `http://dtr-test.pidconsortium.eu/#objects/?query=type:"PID-InfoType"`. The page header includes the site name "Document Repository", navigation links like "Introduction", "All", and "Types", and a user profile for "UlrichSchwardmann" with a "Sign out" button. A search bar is present with the text "Search" and a "Create" dropdown. Below the search bar, a pagination control shows "1 to 10 of 72 results" with a "JSON View" button. The main content area lists several PID-InfoTypes with their descriptions:

- previous-version** [Type: PID-InfoType]
pointer to previous version of digital object by identifier
- next-version** [Type: PID-InfoType]
pointer to next version of digital object by identifier
- ORCID-nameIdentifier-DataCite** [Type: PID-InfoType]
Uniquely identifies an individual or legal entity, according to ORCID schema.
- Other-nameIdentifier-DataCite** [Type: PID-InfoType]
Uniquely identifies an individual or legal entity, according to an arbitrary schema.
- ISNI-nameIdentifier-DataCite** [Type: PID-InfoType]
Uniquely identifies an individual or legal entity, according to ISNI schema.
- time-period-open** [Type: PID-InfoType]
describes a time period between given begin and end date-times. Allow a missing begin or end entry, because it does not allow abbreviated form. A missing begin or end entry in the array means every time before end-time or after begin-time. An empty time period (no date-times specified) represents every time.
- false** [Type: PID-InfoType]
Type Name: caseSensitiveHeader-CSVDDF
indicates that case in the header is meaningful. For example, columns CAT and Cat should not be equated. Default = false
- \r\n** [Type: PID-InfoType]
Type Name: lineTerminator-CSVDDF
specifies the character sequence which should terminate rows. Default = '\r\n'

JSON-Struktur von PID-InfoTypes

DINI-nestor
Session 2

Ulrich
Schwardmann

- `name` : der Name des Typs
- `description` : die Beschreibung des Typs
- `identifizier` : der PID des Typs
- `provenance` : Herkunft, Erzeuger, Zeiten etc.
- `properties` : die verwendeten Untertypen
 - `representationsAndSemantics` : die Eigenschaften verwendeten Untertypen
- `representationsAndSemantics` : die Eigenschaften des Typs
- `validationSchema` : das Schema des Typs
- `standards` : für den Typ relevante Standards

Einführung
Technik

Allgemein
Auflösung
(Resolution)
Management

Granularität

Templates oder
Fragment
Identifizier

Types

PID Information
Types
Data Type
Registries