Leibniz-Institut für
Astrophysik Potsdam

DOI –
datacenters should provide

Harry Enke
Leibniz-Institute for Astrophysics Potsdam (AIP)

Intro: DOI

DOI – digital object identifier
(started in 1998)
- well known entities for scientific papers
- only scarcely deployed for scientific data until recently

History:
- digital identifiers:
  * MAC addresses for network hardware
    and a plethora of industrial id's (=> bar codes)

  * DNS and IP addresses

- analog predecessors :
  * ISBN and other identifier for books etc.
  * catalog systems for libraries

DOI

- notion of a persistent identifier (PI) for digital entities led to
  * various handle systems (Handle, PURL, ARK …)
  * one variant is DOI

".. International DOI Foundation (IDF), [is] a not-for-profit membership organization that is the governance and management body for the federation of Registration Agencies providing Digital Object Identifier (DOI) services and registration, and is the registration authority for the ISO standard (ISO 26324) for the DOI system. The DOI system provides a technical and social infrastructure for the registration and use of persistent interoperable identifiers, called DOIs, for use on digital networks. (www.doi.org)

- addressing major problems:
  * digital objects are by nature volatile, not bound to any real location or physical realisation
  * moving of a digital object leads to difficulties of retrieving finding and verifying it again (link rot)
  * changing references to such digital objects are expensive and should be avoided
- adoption of PIds first in context of librarian efforts to cope with digitised entities, thus the well known DOI applications for publications.

DOI – DataCite

DataCite was founded in 2009, European and US Libraries

* goal: extending DOI to scientific data sets

* registering with DataCite incurs fee (moderate)
  (but: e.g. in Germany academic organisations don't pay)
* contract between organisation and DataCite
* the organisation gets its own DOI prefix

By joining a contract with DataCite the organisation commits to

* guarantee the validity of its DOI
* update the DataCite registry in time when digital objects change their addresses
* objects with DOI should be stable

DataCite

* guarantees resolving of the DOI to the actual address of the object
* keeps a basic set of metadata for each data set

DOI – How it works

We want to publish a set of tables of a cosmological simulation
Example: SMDPL (Small MultiDark Planck) simulation

landing page 1:
- explanation of cosmological parameters, setup of simulation
- URL: https://www.cosmosim.org/simulations/smdpl
- DOI: doi:10.17876/cosmosim/smdpl

landing page 1:
- Description of Rockstar Halo Catalo Table
- URL: https://www.cosmosim.org/simulations/smdpl/smdpl-rockstar
- DOI: doi:10.17876/cosmosim/smdpl/001

landing page 2:
- Description of FoF-Table
- URL: https://www.cosmosim.org/simulations/smdpl/smdpl-fof
- DOI: doi:10.17876/cosmosim/smdpl/002

DOI prefix : 10.17876 ⇔ AIP

DOI – How it works

https://www.cosmosim.org/cms/simulations/smdpl/

Apps Science

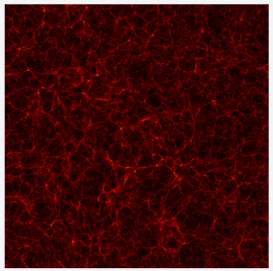
CosmoSim Blog Simulations Documentation Query Admin

SMDPL

The Small MultiDark Planck simulation belongs to the series of MultiDark simulations with Planck cosmology. Its simulation box has a side length of only 400 Mpc/h, less than half of the MDPL-size. With 3840^3 particles inside this small box it achieves a mass resolution of 10^8 M_{sun}/h per dark matter particle.
Some more details for this simulation are described in Klypin et al. 2014.

Please cite this data set using the unique digital object identifier doi:10.17876/cosmosim/smdpl/.

Please give proper Credits when using data from this simulation.



A slice of the SMDPL simulation

Details

Box size	400 Mpc/h	side length of the cosmological cube
Number of particles	3840^3	total number of dark matter particles
Mass resolution	$9.63*10^7$ M_{sun}/h	mass of one dark matter particle
Force resolution	1.5 kpc/h	physical force resolution
Initial redshift	120	redshift at which the simulation started
Cosmology		
h	0.6777	Hubble parameter

gical simulation
ulation

tup of simulation
s/smdpl

s/smdpl/smdpl-rockstar

s/smdpl/smdpl-fof

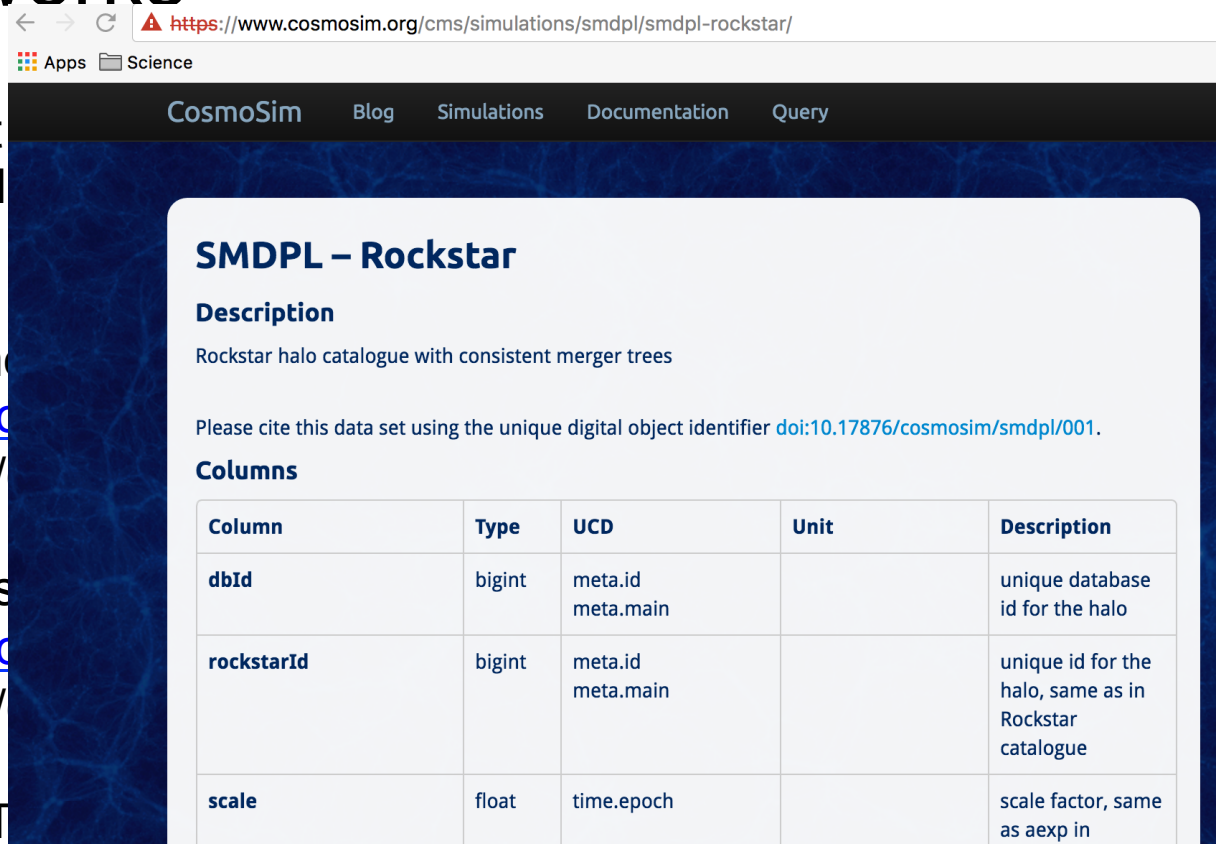
DOI prefix : 10.17876 ⇔ AIP

DOI – How it works

We want to publish a set
Example: SMDPL (Small

landing page 1:
explanation of cosm
URL: https://www.cc
DOI: doi:10.17876/
landing page 1:
Description of Rocks
URL: https://www.cc
DOI: doi:10.17876/
landing page 2:
Description of FoF-T
URL: https://www.cosmosim.org/simulations/smdpl/smdpl-fof
DOI: doi:10.17876/cosmosim/smdpl/002

DOI prefix : 10.17876 ⇔ AIP



DOI – How it works

Required:

* for each data set a metadata file in xml-format
* the website with the landing page carries the doi

Example: rockstar table, doi:10.17876/cosmosim/smdpl-rockstar

prefix

data set location

Upload of metadata:

* via webinterface for single data sets

* via api of DataCite for many data sets
  (but still: call to api for each single doi/data set)

Changes in metadata are versioned by DataCite

DOI –

Required:

* for e
* the v

Exam

Upload of

* via v
* via a
(

Changes i

Dataset: 10.17876/COSMOSIM/SMDPL/001

Metadata Version: 0

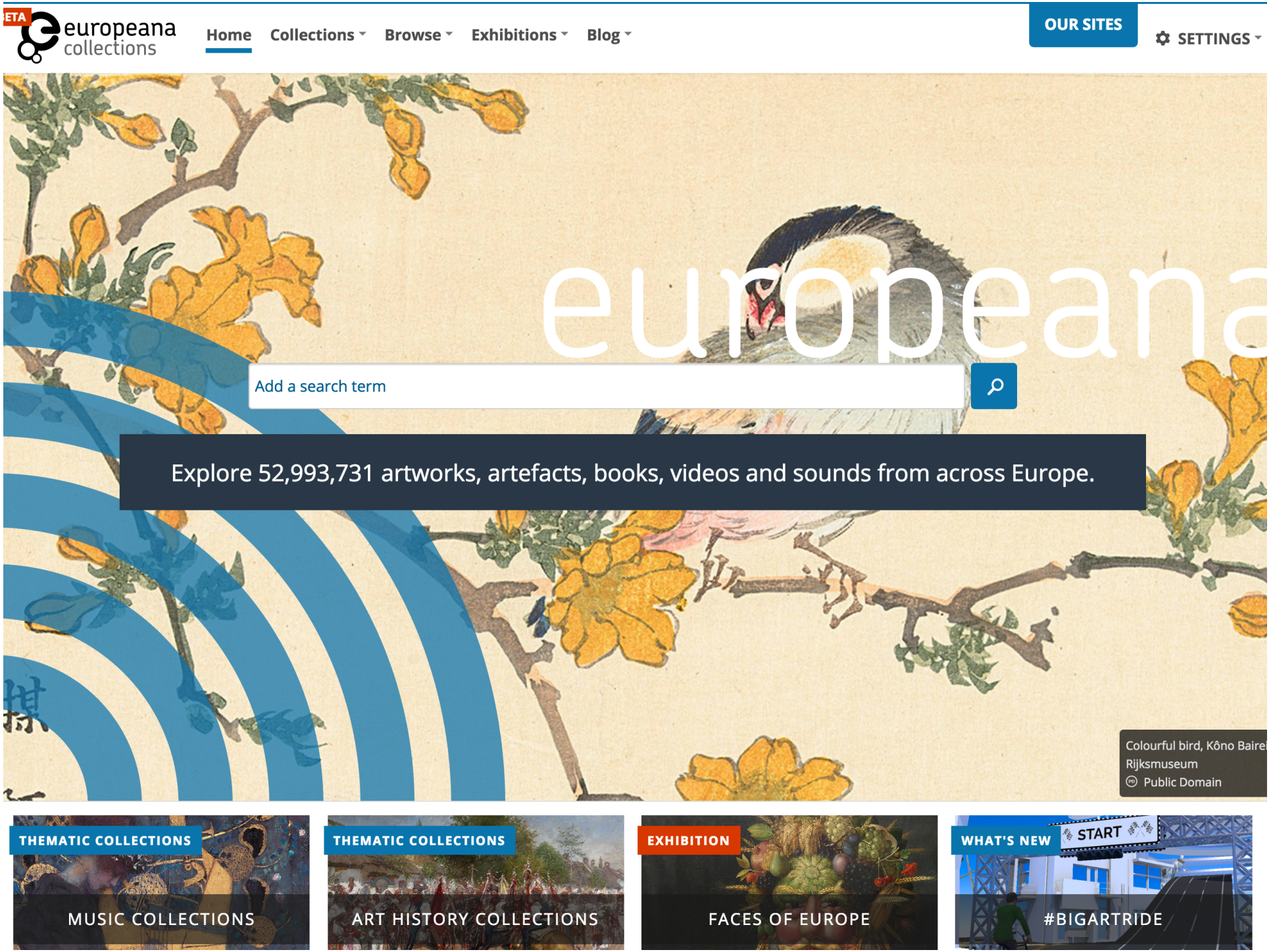
Created: 2016-06-13 20:34 UTC

XML:

```
<?xml version="1.0" encoding="UTF-8"?>

<resource xmlns="http://datacite.org/schema/kernel-3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://datacite.org/schema/kernel-3 http://schema.datacite.org/meta/kernel-
3/metadata.xsd">
  <identifier identifierType="DOI">10.17876/cosmosim/smdpl/001</identifier>
  <creators>
    <creator>
      <creatorName>Cosmosim Database / MultiDark Collaboration</creatorName>
    </creator>
  </creators>
  <titles>
    <title>SMDPL (Small MultiDark Planck) Simulation</title>
  </titles>
  <publisher>Leibniz Institut für Astrophysik Potsdam (AIP)</publisher>
  <publicationYear>2015</publicationYear>
  <subjects>
    <subject subjectScheme="DDC">520 Astronomy</subject>
    <subject>Cosmological Simulation based on Planck cosmology</subject>
  </subjects>
  <contributors>
    <contributor contributorType="DataCollector">
      <contributorName>MultiDark Consolider Project (Spain)</contributorName>
    </contributor>
    <contributor contributorType="DataCollector">
      <contributorName>Leibniz Institut für Astrophysik Potsdam (AIP)</contributorName>
    </contributor>
  </contributors>
  <language>EN</language>
  <resourceType resourceTypeGeneral="Text">Database</resourceType>
  <relatedIdentifiers>
    <relatedIdentifier relatedIdentifierType="URL" relationType="HasMetadata" relatedMetadataScheme="UCD1+"
schemeURI="http://www.ivoa.net/documents/latest/UCDlist.html"
schemeType="Text">https://www.cosmosim.org/cms/simulations/smdpl//smdpl-rockstar</relatedIdentifier>
    <relatedIdentifier relatedIdentifierType="URL" relationType="HasMetadata" relatedMetadataScheme="Units
in the VO" schemeURI="http://ivoa.net/documents/VOUnits/20140523/VOUnits-REC-1.0-20140523.pdf"
schemeType="Text">https://www.cosmosim.org/cms/simulations/smdpl//smdpl-rockstar</relatedIdentifier>
  </relatedIdentifiers>
  <sizes>
    <size>9320923500 rows</size>
  </sizes>
  <formats>
    <format>csv</format>
    <format>votable</format>
  </formats>
  <version>1</version>
  <rightsList>
    <rights rightsURI="http://creativecommons.org/licenses/publicdomain/zero/1.0/">Creative Commons CC0 for
data</rights>
    <rights rightsURI="http://creativecommons.org/licenses/by/4.0/">Creative Commons CC-BY for
images</rights>
  </rightsList>
  <descriptions>
    <description descriptionType="Abstract">Rockstar halo catalogue with consistent merger
trees</description>
  </descriptions>
</resource>
```

discovery metadata


ETA
europeana collections
Home
Collections
Browse
Exhibitions
Blog
OUR SITES
SETTINGS
europeana
Add a search term
Explore 52,993,731 artworks, artefacts, books, videos and sounds from across Europe.
Colourful bird, Kôno Baire
Rijksmuseum
Public Domain
THEMATIC COLLECTIONS
MUSIC COLLECTIONS
THEMATIC COLLECTIONS
ART HISTORY COLLECTIONS
EXHIBITION
FACES OF EUROPE
WHAT'S NEW
START
#BIGARTRIDE

DOI – Europeana

www.europeana.eu

DOI are also applicable identifiers for cultural heritage objects (CHO)

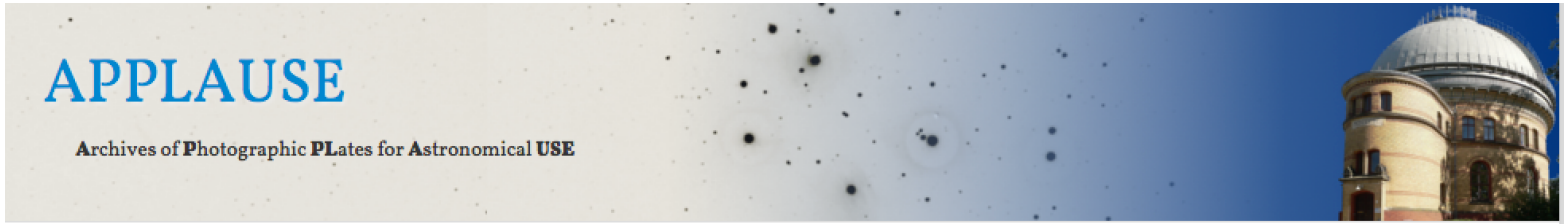
Europeana is a European initiative for publishing CHO
- needs metadata in EDM format
- offers OAI-PMH api for uploads
- has member organisations in many European countries
  - in Germany: collaboration of major libraries (Deutsche Digitale Bibliothek)
- requires contract with organisation
- requires CC0 licensed CHO

Example:

APPLAUSE plate database: ~55000 CHO entries (DR2, 02/2016)

to manage, we use table with metadata and an aid{archive id}
to cope with complex relations between CHO

DOI – Europeana



Project Archives Documentation Query Contact Wiki Login

Objects in Plate Archive

Select object using AID or DOI

Objects in Plate Archive are sorted according to their types as **plates**, **logbooks**, **notes** and **envelopes**.
Each object has an unique AID (Applause IDentifier) stored in corresponding table column as shown below.
DOI (Digital Object Identifier) can be constructed in the following way:

`doi:10.1876/plate/{AID}`

Type	Table	AID column	example AID	example DOI
Plates	plate	plate_aid	dr.2/plates/101_3309	doi:10.1876/plate/dr.2/plates/101_3309
Logbooks	logbook	logbook_aid	dr.2/logbooks/101_53	doi:10.1876/plate/dr.2/logbooks/101_53
Observer notes	logbook	logbook_aid	dr.2/notes/101_12	doi:10.1876/plate/dr.2/notes/101_12
Envelopes	logpage	logpage_aid	dr.2/envelopes/101_8092	doi:10.1876/plate/dr.2/envelopes/101_8092

Landing page of each object can be accessed through DOI or by using of the following form.
Just input the AID or DOI of an object and choose the viewing format.

object AID or DOI | Show object

Web preview EDM XML

DOI – in data centers

* data centers publish data sets
  - which have undergone a quality check
  - which have a set of metadata anyway

* data centers
  - generally have policies for data
  - licenses for usage of published data
  - can guarantee stability for doi mappings

* data centers can provide DOI easily
  - some initial work required
  * create templates for their data sets
  * organise collection of metadata for their DOI
  * have landing pages for each data set with DOI

* data centers can provide a major service
to the scientific community at very low cost

DOI – Use in Virtual Observatory ?

* discussion in Germany already ongoing for some years, no real resolution yet

* no provision (as yet) for special tag in VO table schema

* VO registry asks for services on data, not for the data (resources)
  - could be also one additional field

* VO should incorporate DOI, because
  * science cares for the data, not for the service
  * scientists need the identification of data sets they use,
    - preferably not by indirection
      (query statement + DACHS by TAP service)
    - but by direkt link to data set
      (query statement + DOI)

DOI can connect astronomical data sets to data of the whole science community, not only within astronomy