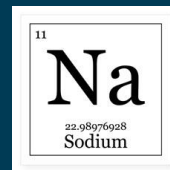


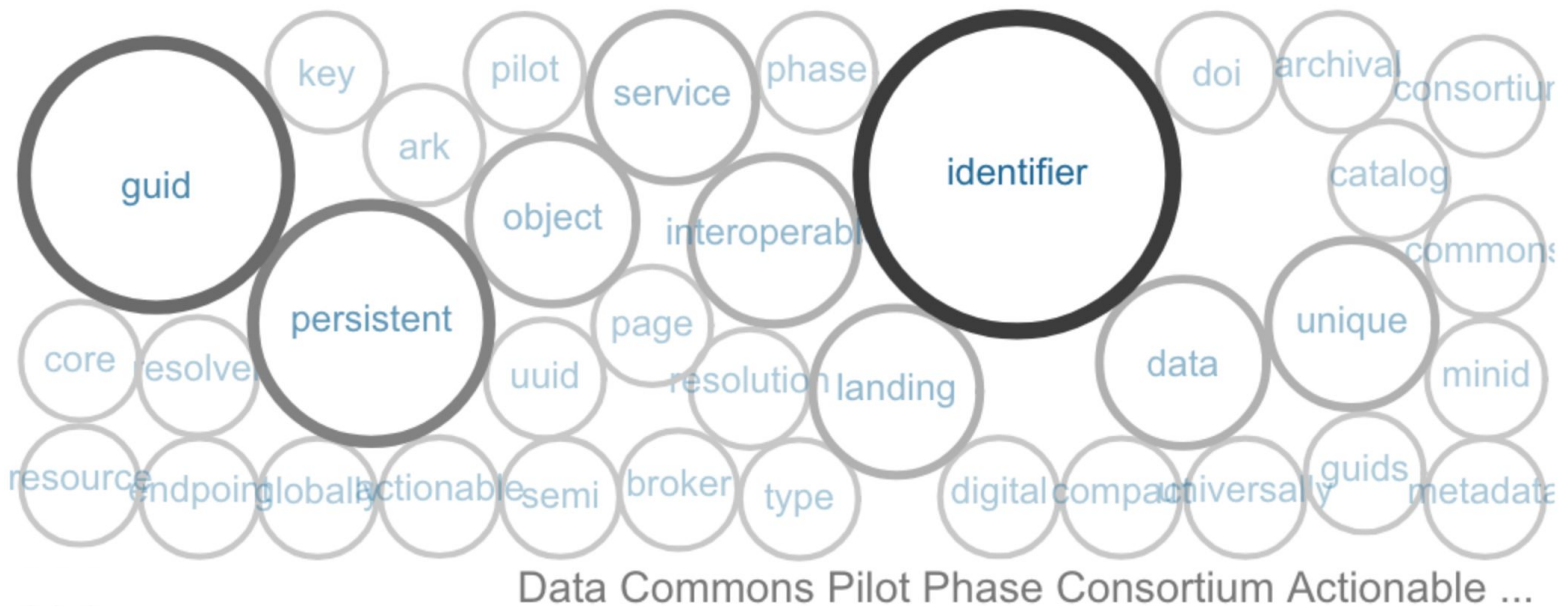
GUID Overview

An Introduction to DCPPC GUID types
and GUID services

Sarala Wimalaratne (EMBL-EBI)



This is a complicated space....



identifier
guid
persistent
data
interoperable
landing
object
service
unique
commons
pilot
phase
consortium
actionable
archival
resource
key
ark
compact
core
metadata

We are here to help:

- 1) GUID Overview
- 2) DataCite DOIs
- 3) Compact Identifiers
- 4) ARKs and Minids
- 5) Data GUIDs

What is a GUID

A **Globally Unique Identifier (GUID)** –

An identifier that follows certain conventions to make it unique within a global context.

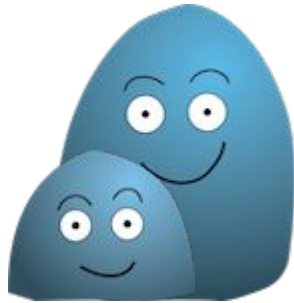
In DCPPC, GUIDs are meant to be **globally unique**, **persistent** within a timeframe, and **actionable** on the Web when prefixed by a resolver URI.



Meet the DCPPC GUID cast



DataCite DOI



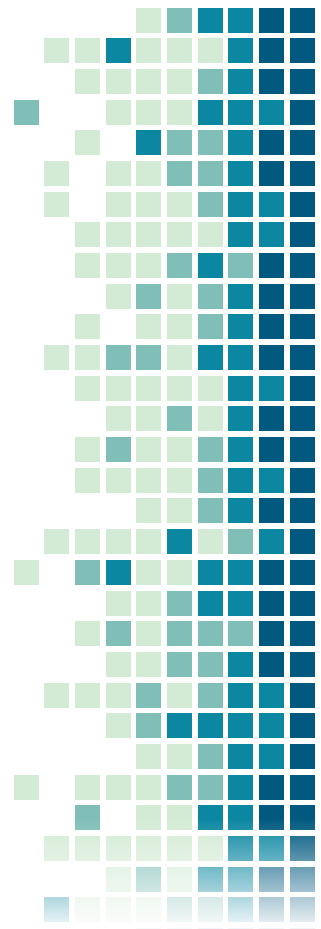
Minid
ARK



Data GUID



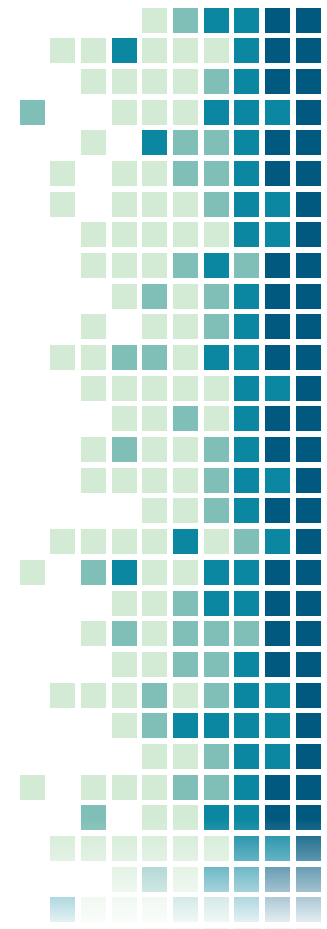
Compact
Identifier



A webinar for each GUID type will follow this webinar

Why GUIDs

- Assigning a **Globally Unique Identifier (GUID)** to a digital object enables you to: 1) locate the object, 2) access it, 3) track it, and 4) connect it to other digital objects across data platforms
- “**Globally Unique**” because we want the identifier to be unique not only within your repository, but across **all** possible repositories



Connecting GUIDs to a Services Model

- KC2 Globally Unique Identifier Services document specifies a commons services model for GUIDs.
- Developed by joint KC2 & Full Stacks team.
- Starting point for further convergence and integration of services and development of common APIs for all GUID types in the next Data Commons Pilot Phase.

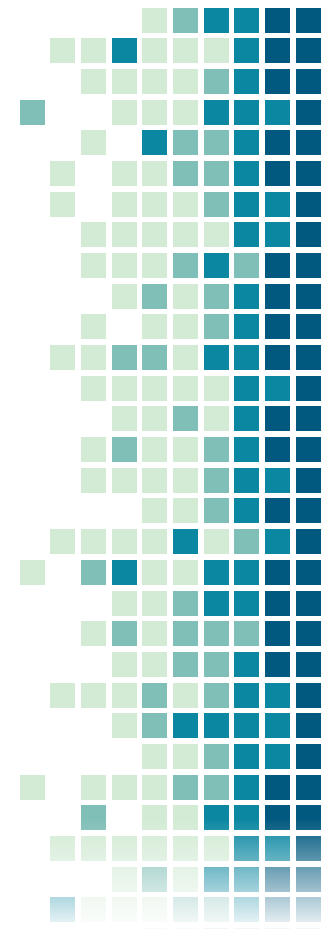
A DCPPC-compliant GUID service conforms to the Globally Unique Identifier Services specification.



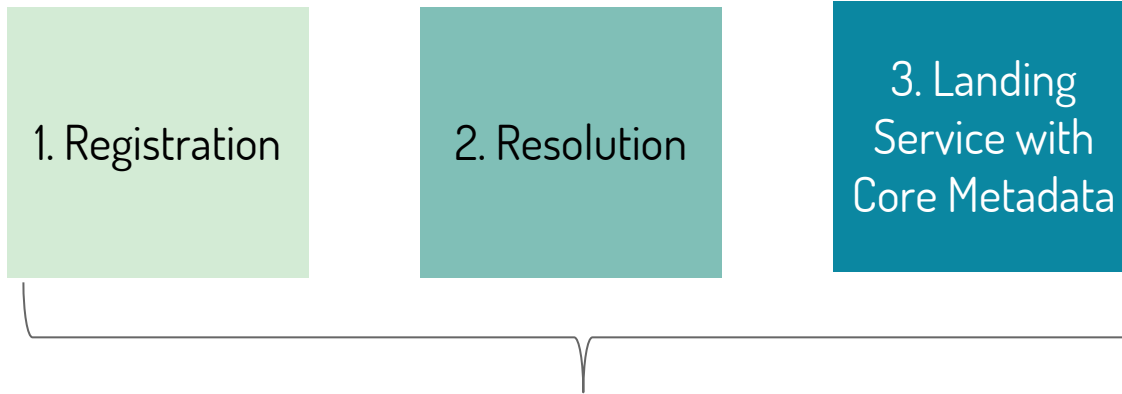
What content gets a GUID

- A GUID can be assigned to an **entire dataset**, to **each file** in a dataset, or even to **data units** within a file.
- A GUID can be assigned to other digital objects (e.g. software, workflows) or to physical objects (e.g. people, reagents).

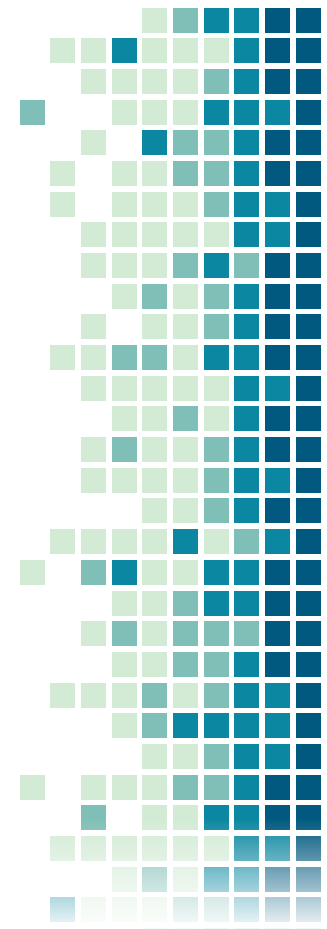
In DCPPC Phase 1, we have prioritized assigning GUIDs to datasets or individual files vs GUIDs for software and workflows.



What does a GUID need to be DCPPC-compliant

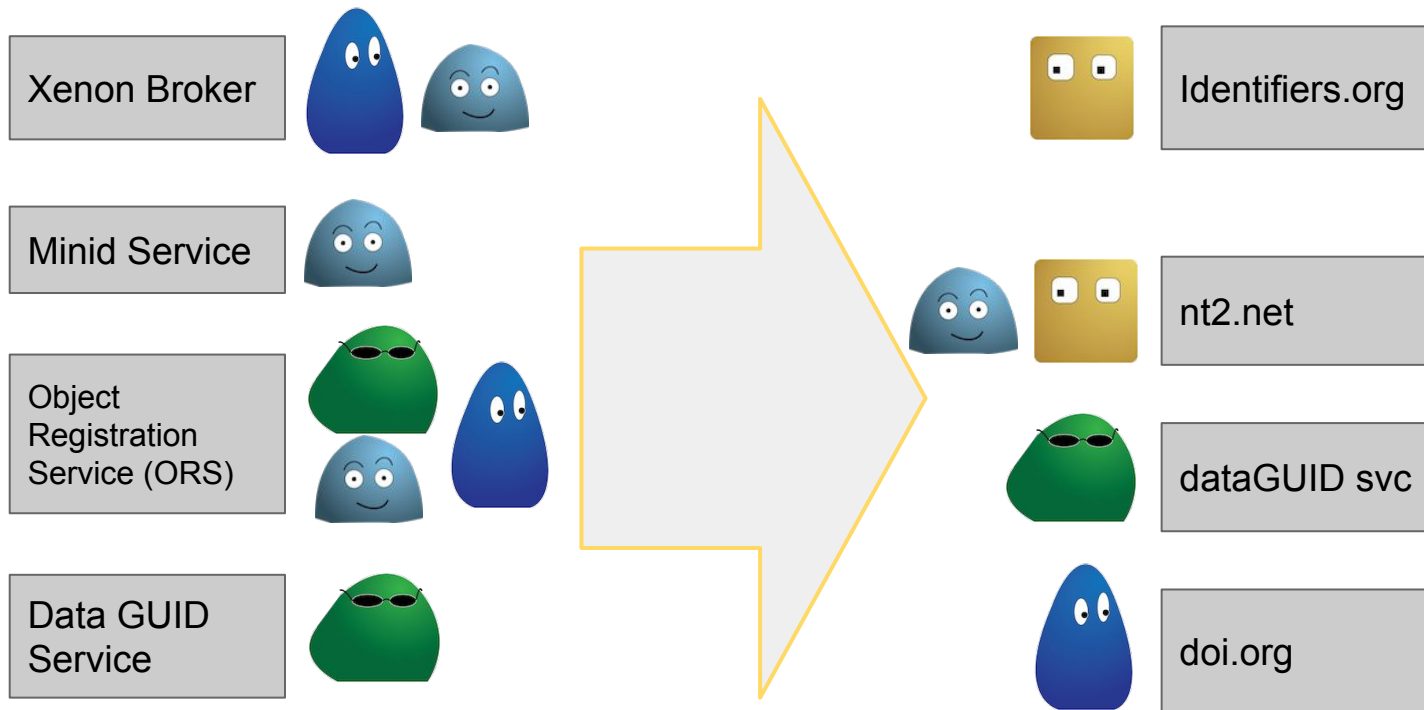


These 3 components enable GUIDs to be **persistent** (within a timeframe) and to **interoperate** across data systems



DCPPC GUID Registration and Landing Services

DCPPC GUID Trusted Resolution Services



= **DataCite DOI**

= **Compact Identifier**

= **Minid /ARK**

= **Data GUID**

Registration

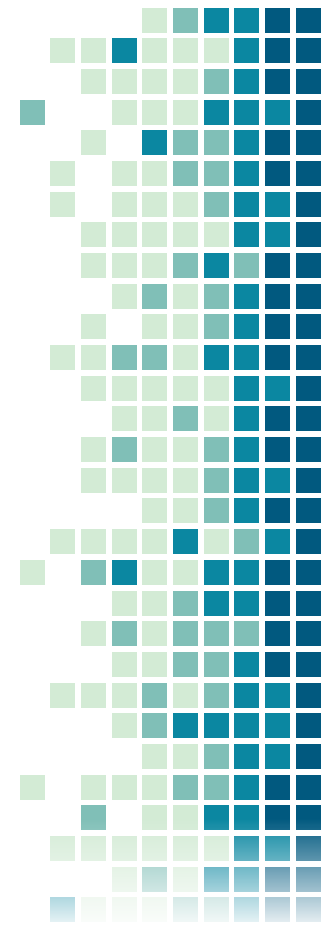
- A GUID must be registered in a global registry with core metadata, to guarantee uniqueness, findability, accessibility, and verification
- Two types of registration:
 1. Namespace registration: identifiers.org
 2. Object registration: ORS, doi.org, n2t.net...



Resolution

- A resolver service guarantees that a GUID will resolve to the same object, even when the location of the object changes
- GUIDs must resolve to a landing service or page, not directly to the digital object

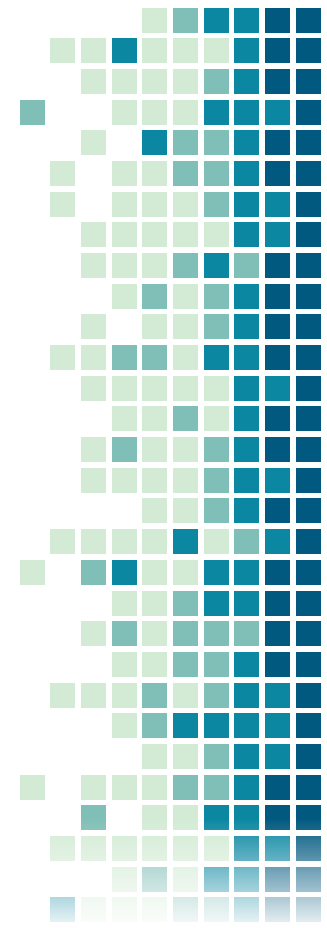
DCPPC resolvers: identifiers.org (Compact IDs); n2t.net (ARKs / minids & Compact IDs); DOI.org (DOIs); DataGUID service (dataGUIDs).



Landing Service

- The landing service or page that a GUID resolves to must provide human and machine readable GUID core metadata.
- It can also provide additional metadata.
- It should provide information on how to access the digital object, via one or more cloud services

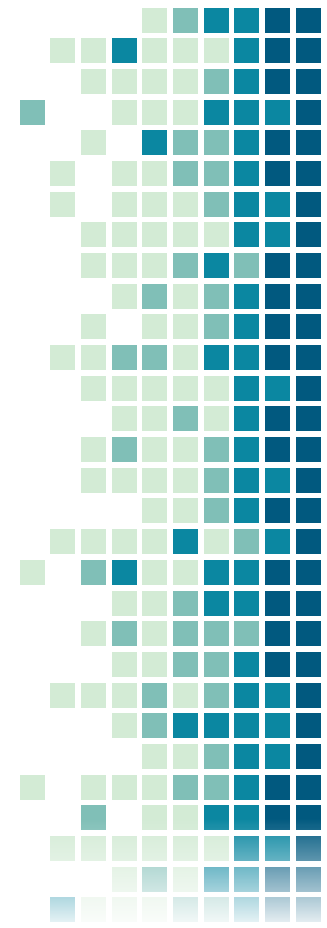
A DCPPC-compliant landing service is one that provides at a minimum, required core metadata for a given GUID type, as schema.org JSON-LD.



GUID Core Metadata

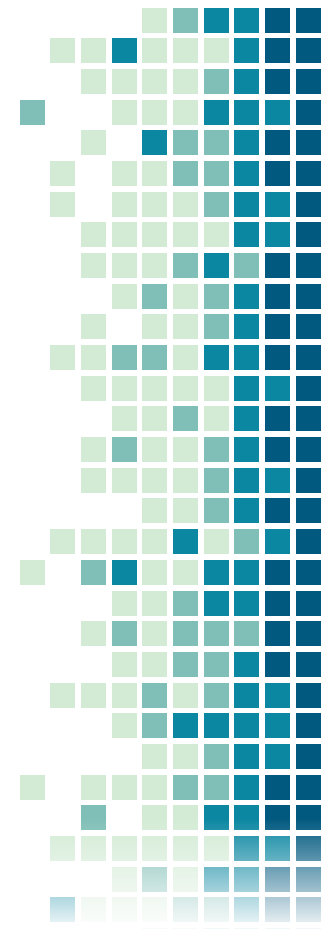
- Core Metadata enables GUIDs to interoperate across systems exchanging descriptive and access metadata
- The core metadata, accessible from the dataset landing page or service, should be human and machine readable

In DCPPC, each GUID type must support a minimal subset of core metadata fields appropriate for that GUID type. Additional fields are optional.



Object Registration Service (ORS)

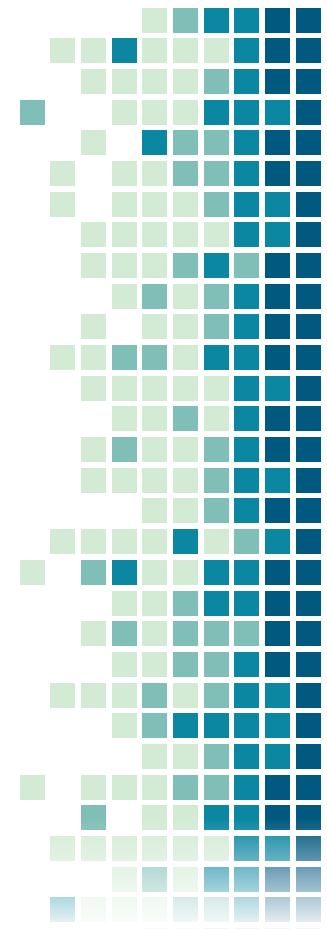
- **ORS** (Team Sodium) provides registration and landing services for DCPPC GUIDs.
- **KC-compliant** with
 - KC2 (CoreMetadata & GUID Services)
 - KC3 (SmartAPIs).
 - KC6 (OAuth2/GlobusAuth).
- **Integrated** with Datacite and CDL global services.



What are the benefits of
DCPPC-compliant GUIDs?

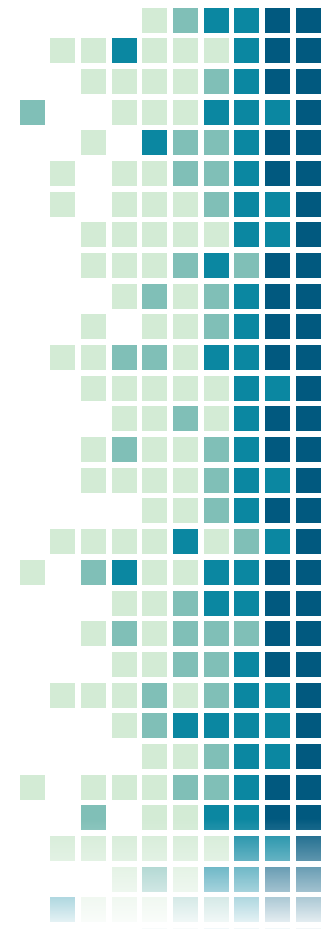
Share Datasets and Compute across Full Stacks

- Common model to register GUIDs and resolve them to a landing service with Core Metadata
- Allows each full stack to use GUIDs registered by other full stack, with a common understanding of what that GUID is



Discoverable by Google Dataset Search

- Embedding schema.org in a landing service/page as required by DCPPC Core Metadata makes every dataset in DCPPC discoverable through Google Dataset Search
- Potentially can add more fields to schema.org/dataset to improve Google Search results



Data Citation

- Enables validation of research results.
- Enables reuse of research results by others.
- Translates to credit for data creators, data stewards, and experimentalists.
- Persistent GUIDs are needed to generate a formal data citation acceptable to publishers, journals, and institutions.



Q&A