

**VRE for regional Interdisciplinary
communities in Southeast Europe and
the Eastern Mediterranean**

**Persistent identifiers in
VI-SEEM**



Petar Jovanovic
Institute of Physics Belgrade

- ❑ What are persistent identifiers?
- ❑ Why are they useful?
- ❑ Different PID systems
- ❑ Structure of the PID
- ❑ PIDs in VI-SEEM
 - ❑ The HANDLE system

What are PIDs? (1)



- ❑ Data generation is getting easier and cheaper.
- ❑ More complexity in data processing and analysis.
- ❑ Both amount and quality of data output are increasing.
- ❑ To stimulate reuse and reproducibility, data needs to be
 - ❑ accessible
 - ❑ reusable
 - ❑ searchable
 - ❑ interoperable

What are PIDs? (1)



- ❑ Data generation is getting easier and cheaper.
- ❑ More complexity in data processing and analysis.
- ❑ Both amount and quality of data output are increasing.
- ❑ To stimulate reuse and reproducibility, data needs to be
 - ❑ accessible
 - ❑ **reusable** ←
 - ❑ **searchable** ←
 - ❑ interoperable

What are PIDs? (2)



- ❑ Pointers to data resources (data files, metadata, documents...)
- ❑ Globally unique
- ❑ Infinite lifespan

- ❑ Can be used to identify and retrieve resources.
- ❑ Can be resolved to a resource.

- ❑ Examples: ISBN, DOI, PURL, Handle...

Why are they useful?



Why not just use URLs?

- they specify a network location for digital resources
- they depend on a server
- they can change
- resources might need to be moved

Over a long enough period it is not uncommon for URLs to stop working.

PIDs are designed to be persistent over time.

Why are they useful? (2)

- ❑ Persistent identifiers are
 - ❑ not URLs
 - ❑ not strictly bound to a network address or file names
- ❑ Wikipedia:

“A persistent identifier (PID) is a long lasting reference to a digital object -- a single file or set of files.”
(https://en.wikipedia.org/wiki/Persistent_identifier)
- ❑ Identifier just points to a resource without actually assuming anything else about the resource.
- ❑ PID owner is responsible to keep it up to date when the resource changes.



Structure of PID



- ❑ **Prefix** designates administrative domain, comes from an issuing instance.
- ❑ **Suffix** is unique in the realm of the prefix.
- ❑ The resource is globally addressable with its PID.
- ❑ PID does not change over time or with resource location.
 - ❑ The process of resolution encapsulates the changes if they happen.

- ❑ **Persistent URLs (PURLs)**

purl: XKC/xkc3032

metadata: no additional metadata

- ❑ **Archival Resource Key (ARK)**

ark: /23321/32423xk321

metadata: Electronic Resource Citation

- ❑ **Digital Object Identifier (DOI)**

DOI: 10.1000/192

metadata: the INDECS schema stored in a separate database

PID system requirements



- Attach multiple URLs to a PID
- Allow part identifiers for complex objects. (granularity issue)
- Allow attaching of extra metadata to the PID
- Actionable PIDs (i.e. convertible to URL)
- HTTP proxy for resolving (port 80 only)
- Controlled by community
- API for administration of PIDs from applications
- Delegation of PID administration to other organizations
- Distributed, robust, highly available, scalable
- No single point of failure, distributed system with mirroring
- Acceptable non-commercial business model

Identifier string requirements

- ❑ Not based on any changeable attributes of the entity (e.g. location, ownership...)
- ❑ Unique
- ❑ Opaque, preferably a number
 - ❑ a well known pattern invites assumptions that may be misleading
 - ❑ meaningful semantics invite problems
- ❑ Nice to have:
 - ❑ human-readable
 - ❑ cut/paste supported
 - ❑ fits common systems (like the URL specification)

- ❑ Based on HANDLE system
- ❑ Example:
 - ❑ PID: 21.15102/VISEEM-276
 - ❑ Actionable PID: <http://hdl.handle.net/21.15102/VISEEM-276>

Handle.Net[®]

Handle Values for: 21.15102/VISEEM-276

Index	Type	Timestamp	Data
1	url	2017-10-11 12:44:14Z	https://repo.vi-seem.eu/handle/21.15102/VISEEM-276
100	HS_ADMIN	2017-10-11 12:44:14Z	handle=21.15102/REPO; index=301; [create hdl,delete hdl,read val,modify val,del val,add val,modify admin,del admin,add admin]

- ❑ Used in VI-SEEM data repository and other services.

- ❑ Persistent identifiers provide a solution to “link rot” by providing an extra layer of indirection.
- ❑ Several systems are available.
- ❑ The HANDLE system is the foundation of VI-SEEM PIDs

- ❑ Persistent identifiers provide a solution to “link rot” by providing an extra layer of indirection.
- ❑ Several systems are available.
- ❑ The HANDLE system is the foundation of VI-SEEM PIDs

Thank you for your attention.