Structural Elements in a Persistent Identifier Infrastructure and Resulting Benefits for the Earth Science Community

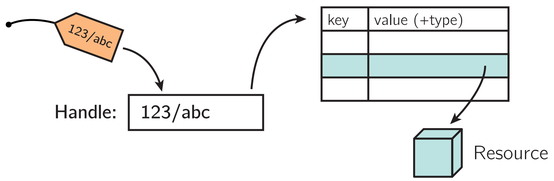
AGU 2012; IN23C-1524; Control ID: 1464410

T. Weigel[1,2], F. Toussaint[1], M. Stockhause[1], H. Höck[1], S. Kindermann[1], M. Lautenschlager[1] and T. Ludwig[1,2]
E-Mail: {lastname}@dkrz.de
[1]Deutsches Klimarechenzentrum, [2]Universität Hamburg

Handles as defined by Kahn and Wilensky (2006) are a specific type of Persistent Identifiers.
Handles have one purpose only:

Handles resolve to key-value pairs.

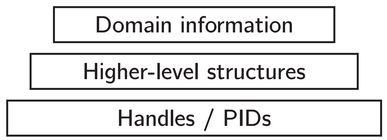
These key-value pair are also called key-metadata.
Typically, one of the values points to a resource URL.
All Handles form one continuous single namespace.
The referenced resource can often be viewed as a black box.
Handles are cheap. We can create millions of them.



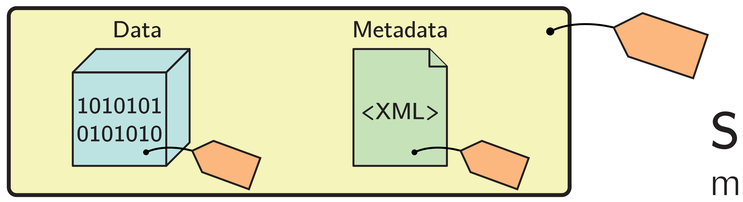
A user enters a Handle in a resolver interface and will be redirected to the resource. Tools can also access the key-metadata to provide higher-level services.
The best example: DOIs. All DOIs are Handles.

The idea: Use key-metadata to encode **typed links** between Handles!

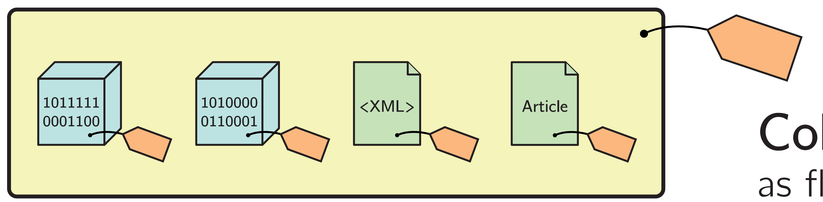
PIDs enable a layer-based architecture. Each layer can support software tools that do not have to know about the contents of upper layers.
We can use the same tools to manage e.g. replication of language resources as well as Earth science resources.



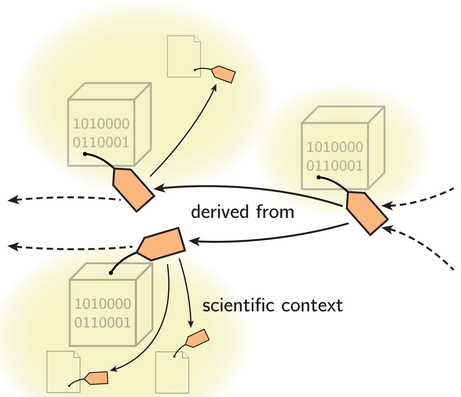
References:
1. Kahn and Wilensky (2006): A framework for distributed digital object services. Int. J. of Digit. Libr., Vol. 6, No. 2. doi:10.1007/s00799-005-0128-x
2. Duerr, Downs, Tilmes et al. (2011): On the utility of identification schemes for digital earth science data: an assessment and recommendations. Earth Sci. Inf., Vol. 4, No. 3. doi:10.1007/s12145-011-0083-6
3. Weigel, Lautenschlager, Toussaint, Kindermann (2012): A framework for extended persistent identification of scientific assets. (under review)



Small sets
millions of them!



Collections
as flexible as the use cases



Provenance Digraph
with typed edges

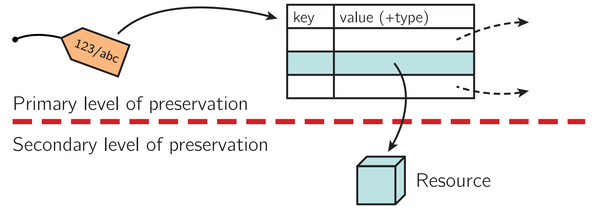
The Research Data Alliance

Candidate Working Groups include
- PID Information Types
- Information Type Registry
- Metadata
- Data Foundation and Interoperability

... and many more!

Join us at the **official RDA Launch**
March 18-20 2013 in Göteborg, Sweden.

http://www.rd-alliance.org



Long-term archival can benefit from distinguishing two **layers of preservation**.

The imperative: Put structural information in most persistent layer.

1. This preserves scientific context. There might be some value to keeping PIDs even if the original data is lost (Duerr et al., 2011)
2. This keeps structures such as the provenance graph intact even if data is deleted (such as intermediate results).
3. It enables automated data management tasks such as replication. The replicating system does not need to know the kind of elements it is dealing with, but it must know about their structural dependencies.

There are examples for the use of collections in existing Handle-based systems: DataCite DOIs of a dataset collection or IGSN parent-child structures. Currently these are maintained in external databases. Using our proposed structures, they may be encoded at the more persistent layer, enabling agnostic tool chains.

If used widely, these structures can enable solid navigation from a DOI to not only relevant data, but also to the elaborate context of a scientific work. This is very straightforward **because DOIs and IGSNs are technically based on Handles!**

The **Research Data Alliance** is an emerging forum to channel these activities and build consensus on practical use.

There are fundamental requirements to be considered.

1. Identifiers must be globally unique
2. Identifiers must be globally discoverable
3. Resolver services do not behave differently from each other
4. The resource link must be modifiable
5. Resolution operations always return the same results over time
6. Key-metadata remains available even if the resource is gone

More details: Weigel et al., 2012, forthcoming

http://www.dkrz.de
http://www.wdc-climate.de

DKRZ
DEUTSCHES KLIMARECHENZENTRUM