NOAA Talking Points

Slide 1
Good Afternoon. I am happy to be speaking with you today about Data Publication and Citation and the use of DOIs. I wanted to share these pictures – these are some of the people who have been working on this issue of data publication for several years.

Slide 2
We are a group of scientists, data managers and librarians that came together and shared our expertise. We felt it was important that all these groups participate and contribute to address the data publication paradigm.

Slide 3
Our challenge was to develop a system that would support citation thus increasing the availability and access to research data.
The benefit to the researcher would be proper attribution and recognition of the importance of data curation. As Peter Fox says – treating data as a first class citizen.

Slide 4
Our Project was based on two use cases. One – data related to journal articles are deposited into an institutional repository and assigned a persistent identifier. Use Case two is data held by data centers are packaged and served in formats that can be cited.

Slide 5
The goal of the use cases has been to indentify best practices for tracking data provenance and attribution. The assignment of persistent identifiers, in this case, DOIs enable accurate data citation.

Slide 6
This is a workflow diagram. Our intention is to show that fieldwork to citation for data can follow the same path as publication of a paper. The end result for both being a citable DOI.

Slide 7
Data publication that enables accurate data citation can be an incentive for researchers to share their data. Data citation metrics are starting to be collected.

The question is will they be considered in performance assessments … and of course some of this is being driven by funding agency and publisher mandates.

Slide 8
This slide will be simplistic for some of you, but I did want to include it for those less familiar with the specifics of DOIs. Librarians have been using DOIs for years and they are now becoming the de facto standard for data. Many current data projects register their DOIs with DataCite, an organization that is working to develop standards that enable data access and reuse. The MBLWHOI Library registers our DOIs with CrossRef. We had been assigning DOIs before there was a DataCite and the relationship has continued to work well for us. Many of the major publishers use CrossRef.

Slide 9
Back to our use cases. Use case one deals with data associated with journal articles. In Woods Hole we are working directly with authors as well as data managers to deposit data and metadata. In the example here an author came to us before he submitted his paper. We deposited the data into our Institutional Repository – the Woods Hole Open Access server and assigned a DOI that he included in the article. More commonly we are depositing data and assigning DOIs after publication of an article.

Slide 10
In either case the metadata record for the article contains a link to the dataset and the dataset record contains a link to the article.

Some of you may recognize the look of this record format – we have a DSpace repository.

Slide 11
In 2011 Elsevier saw one of our project posters at a conference and approached us about a linking project. Now when there is data in the Woods Hole Open Access Server associated with an Elsevier publication a banner appears in Science Direct pointing to them to the data.
This is a nice added feature for our researchers.

Slide 12
The Library also has a collaboration with the Biological and Chemical Oceanography Data Management Office at WHOI. We have automated ingestion

of datasets and metadata into the Library repository and the system can return a DOI to BCO-DMO. The data manager reports that researchers are very interested in this feature.

Slide 13
Here is a sample of a BCO-DMO record that includes the DOI assigned through the Library.

Slide 14
Use Case 2 focuses on data from a data center. The British Oceanographic Data Centre is selecting datasets, exposing a fixed copy and assigning a DOI obtained from the British Library. This publication process allows datasets to be cited and accessed in the future. You can compare this process to the old days when data centers would publish datasets on CDs or DVDs and distribute them to libraries and data centers.

Slide 15
Here is a sample record from PDL. Notice the DOI is in the metadata record, just as it was in the WHOAS and BCO-DMO examples.

Slide 16
There are other successful similar data publication/citation models including Dryad and Pangea.
This type of repository is not meant to replace national or subject repositories. It is a service we can offer researchers to make their data citeable.

Slide 17
We have published a cook book for those interested in doing data publication at their institutions. It is available at IOC/IODE Manuals and Guides – Number 64.

Slide 18
This is not difficult. Even small organizations can implement a data repository of this type.

Slide 19
This project is ongoing and we will update the cookbook as appropriate. To give you an idea of our forward direction ……

A next step for us is to implement ORCID IDs. The library had been monitoring the progress of several researcher id projects before we made the decision to support ORCID and encourage our researchers to register.

Slide 20
Our decision has been based on the strong support of some major players, as well as interoperability with other ids and systems. Upon the recommendation of the Library it looks like people we collaborate with like BCO-DMO and the R2R project will also adopt ORCID ids.

We are trying to stay flexible and open to the new possibilities of linking repository data and text with other resources.

This is a work in progress.

Slide 21
Thank you.