

## Questions asked during the May 2013 Deep-dive with DPN Members:

- *Do you think of DPN as providing higher availability and reliable?*

- A: Yes, but in the sense that it's sitting behind the front-facing archives, but is itself a dark archive. The first nodes are the connections to data owners and provide the front-facing services.

- *Legal requirements: our institution has requirements for a variety of legal issues. Does DPN have specific issues or requirements in this area? Especially with the new rules coming down from the Whitehouse and other federal agencies.*

- A: DPN will be sitting behind the front-facing archives and buttress their needs, including legal and regulatory issues. And we will be working with funders to make sure we help them meet these requirements.

- *What do you mean by "dark?" Does it mean encryption?*

- A: DPN is a dark archive in these sense that data in it is not considered actively available for access systems. It's data that has been "put away" for safe keeping. Encryption of objects within DPN is possible but not necessary.

- *If we have a range of services, does that mean we might be working with multiple first nodes?*

-A: Possible, yes. Necessary, no. This will be a decision that individual institutions will need to make, based on their local needs.

- *Will all data go to all nodes:*

A: No. We are still figuring out what the best number of replications is.

- *What is the provisioning strategy for DPN storage? How will decisions about where objects go in the network happen?*

A: We are still working on it - no answer yet. This is something the DPN management will need to determine, in relation to the services and business model discussions. Likely it will also revolve around geographic location, political affiliations, etc.

- *Is it assumed that the content held in the current first nodes automatically will be part of DPN, or something negotiated?*

A: TBD. This decision will be part of the business and policy discussion.

- *Is architecture designed for fewer bigger nodes or many thousands of nodes?*

A: Generally fewer, bigger nodes, especially because of the functions tied to the first node discussion. Also note the number of trust and security questions, which points to a smaller number of nodes.

- *Is there any kind of contract that is tied to a first node based on functions? Is there money set aside for handling this?*

A: Right now, no. Contractually the nodes are the same, and dark. Also note

that the discussion has been in terms of "ungraceful" dissolution of nodes. It's also the case the the individual nodes are heterogenous and also have their own succession plans in place that would be outside of DPN. This is something that needs to be added into the policy discussion within DPN. NB also the key management discussion.

*- Is there a possibility that material would be removed from DPN because of some contractual obligation?*

A: One concept is that once material is in the network it would remain in the network. There is a notion that content would leave the network for any number of practical reason, and we are working on the policy details around this.

*- What about content that is already held in repositories that might be put into DPN at this point? Still need to figure that out and harmonize as needed.*

*- What if a donor doesn't pay its bills?*

A: A succession event would get kicked off, based on the contracts in place. The business group is also working to get a single up-front payment that would help ameliorate about this.

*- Will there be an API for reading and writing into the registry?*

A: No. There is a specification for the messages and it's up to the individual DPN nodes to implement the local registry based on them.

*- Doesn't that complicate the registry process?*

A: It's simpler. Because we aren't maintaining a copy of the same registry, but a synchronized implementation of the node at each locale.

*- What about a new replicating node? What happens for them in terms of the registry?*

A: A new node coming along would get a dump of the previous information so that they can populate their own registry.

*- How do we do an audit to find out where a bad entry in the registry comes from?*

A: In the easy case, the first node is doing an audit of its content. It should detect and be able to fix it. We also can go back into the actual bags that we're holding and re-populate the registry as needed.

*- Have you decided on an identity management scheme and how it maps to our local implementations?*

A: Yes, we have a plan in place for how to reconcile local identifiers and DPN identifiers. We definitely want to have a DPN-specific identifier but not burden the first nodes with this. This will be tracked within the registry. It's also not required for a local archive to implement a separate ID outside of the DPN one.

- *Is it necessary for each replicating node to have the "thing" that is referenced in the registry?*

A: It should not matter. You should be able to reference or get it based on the registry entry. Spreading across the network is more desirable in the future as the network grows.

- *Could you put into DPN local references - e.g. a local license applying to my collection?*

A: Yes.

- *Are you going to be keeping stats on the data that's coming in, since it will provide an interesting research corpus?*

A: if the content is dark, do we really want to be doing that, other than what the first nodes are doing? Likely not, but we might want to have some kind of preservation registry of some kind.

- *What about best practices recommendations?*

A: These should come out of the first nodes, since they are the actual archival repositories.

- *Are you building a policy framework around the messaging infrastructure?*

A: Yes, we will have this in place. It's not there yet, and it will need to be nuanced to handle all of the various needs.

- *During the initial data transfer, does the replicating node receive the checksum with the object?*

A: No. They will generate it and then send it back to the first node for verification.

- *What is the timeframe for a fixity check?*

A: Within DPN we have a concept of requesting fixity for items. We haven't finalized the policy of how/when/who.

- *Do we anticipate migrating our checksums when SHA-256 is no longer the standard?*

A: Yes. At this point the estimate is that it will be mid-2030s until it's broken.

- *Will you have audits of registries?*

A: Yes, at the local nodes. If discrepancies are detected, they will be added to a suspect list for future examination .

- *For data deletion, what about datasets that are required to be kept for x numbers of years?*

A: This kind of policy is possible. It's also arguable that maybe it shouldn't have been put in DPN in the first place. But you still want to make sure the DPN is fitting in to the funder requirements. Also, there should be policy and business

work involved here. And again, the first node should be driving this decision making.

- *Do you keep any kind of audit trail for delations?*

A: Not fully answered right now, but still a policy decision that needs to be fully fleshed out.