Sponsor Summit Open Forum Discussion Notes

The notes below reflect a summary of the March 13 open forum discussion answer the proposed questions.

Wednesday, March 13, morning sessions (cmm)

James Hilton kick-off and *presentation*

- It is not a crisis or cliff pushing this meeting agenda. It's an opportunity to come together and "Move the needle" towards a future that meets all of our needs.
- We are at a pivital moment that hangs on (digital) preservation.
- A 2011 Nobel Prize was awarded for those who noticed that the universe is expanding faster and faster; it is not static. In the far future galaxies that we now see will rush away so fast that we won't be visible anymore. The blackness will infer that nothing else ever existed in the universe except that matter which is immediately in front of and surrounding humans in that distant future. It is our responsibility to ensure that evidence of stars, galaxies and solar systems inform the far future.
- We have an emerging **digital preservation stack** (see JH slides).
- Where do all these parts and pieces fit together other than around the 20K price tag for support of each?
- No guarantee that any of these layers feed into the rest of the stack (bleed between the layers); institutions may utilize parts and pieces of the stack:
    - Access repositories
    - Preservation repositories
    - DPN backbone
    - Code (that includes DSpace and Fedora)
- What do I get for my 20K annual fee?
    - It is an investment in collective/shared capabilities; it's more like thinking about FTEs because that's where you get "capability"

Discussion

- Rob Catalano--stack of services is not yet clearly defined; still in R&D phases
- We need to understand the ecosystem that DuraSpace lives in
- Dean Krafft--How do I decide which to invest in?
    - If we fail to invest as a community we will suffer the tragedy of the Commons
    - No reason not to be putting the investment into all the layers but not everyone has to invest
    - We need to get the right mix/compliments among efforts on the stack; what is the most effective investment
    - The stack helps us winnow/select content for preservation

JH Short overview of Fedora past

Institutions and foundations invested heavily in DSpace and Fedora and it was good. Millions of dollars were spent. Institutions and foundations said they were done. "You must be self-sustaining". There was no spigot with millions of dollars coming out of it. How do we make the organization self-sustaining was a DuraSpace board question. How to sustain community open source projects is a community issue. Revenue was diversified over a number of years with the sponsorship program and grants along with service revenue. This is the moment to take up the question sustainability for open source projects.

Michele and Jonathan Myths and Debunks *(see slides)*

- "Our Mission is not about software; we are committed to our digital future." MK
- We are here today because we need to work with you and creating sustainable, new investment strategies to reinvigorate DSpace and Fedora developmen
- Fedora needs about a half million dollars per year to move forward toward rapid, cohesive development
- "Each of you owns the Fedora and DSpace code to the degree in which you participate and utilize the open source software. You all own Fedora and DSpace." JJM
- How does VIVO fit in? (Dean Krafft)
    - VIVO has moved from being a grant-supported project into developing a model for sustainability.
    - Large grant from NIH
    - More than 100 institutions piloting VIVO software
    - Signed on as a DuraSpace incubated project because:
        - Wanted independent biz administration
        - Wanted help with infrastructure (wiki, code repository)
        - Looking for a model for sponsorship and project support
        - There will be a tech coordinator/lead who will be hired through DuraSpace
    - "VIVO fits into the DuraSpace mission because research institutions like Cornell need to know WHAT the university scholarly output is (that should be preserved)--that's the content that VIVO pulls together." DK

Discussion

What are the interrelationships among our efforts?

- There is a difference between stewardship and ownership. The DuraSpace board and organization are charged with the fiduciary responsibility of keeping the org going; communities have to "own" the projects
- "Dude, you own the software."

How is governance going to come out of this somewhat new perspective (for this group)?

- Is some of the ambition of the meeting about moving projects into the services column?
  - Not necessarily. We have done the low-hanging fruit service (DSpaceDirect). We are not in a crisis and are not abandoning DSpace or Fedora--have enough $ for immediate future. Clarifying roles of community and DuraSpace
  - Less about sustainability and more about momentum
  - We cannot continue to fill the gap in development costs
  - Fedora is a piece of code; "FedoraCloud" would be a service
- VIVO sponsorship level is higher than DSpace or Fedora levels; expectations are that the founding sponsors will set the strategic direction for VIVO; looking into a service offering that will help with VIVO sustainability
- Products not projects?? Projects are not backed up with sales of a product. This is how people know how they are doing and what they are doing.
- How many people believed some or all of these myths (early slide) before this presentation (mixed show of hands)? If people leave understanding this slide the meeting will be a success. This meeting is really about the future of Fedora and DSpace.
- How can we engage the other 93% of repository users (free riders) in new ways?? Can it be done either financially or with contributions in kind?
- Have to find ways for folks outside of ARL to contribute, perhaps non-code, testing, usability perhaps. Many institutions are not able to contribute at the 2-FTE level
- Lifecycle for a stack of software is 5-10 years. This is a different timing issue that general preservation policy and scholarship has already gone digital. "We are in humanities in the digital age."

Tyler Walters (see slides)

- VT participation $122,904 total = 1.4 FTE
- ARL library; 17.2M dollars annually
- For this annual investment we get many benefits
- Where do the larger increments of dollars that it will cost to participate in DPN beyond 20K--of course there are ongoing costs for having staff participate; leverage the scarce developer resources we have with those of other shops
- When talking digital preservation the language is often not understandable by those people who fund preservation efforts
- We now enough digital content that u administrators value that we can start having the DPN conversation

Discussion

Does DPN need DuraSpace?

- Code layer at the bottom is necessary to manage your digital content in a way that is visible and meaningful for scholars and scholarship. We need DSpace and Fedora to do this.
- Without the stack there ain't no DPN; DPN needs repositories. Is the value proposition worth it to the DSpace and Fedora communities?

Does Tyler have an idea of how much that $$ would buy without the community efforts and projects?

- The way it works for Virginia Tech is idiosyncratic and involves a lot of "pushing around"

Sayeed, We need to put a timeframe around this. There is an urgency. Data is being destroyed right now. It is being destroyed now--what was wrong with those people (us).

- There is not any chance to preserve some materials that are so fragile they are deteriorating.
- Invisible to visible--Ann Wolpert. "You can't see digital."
- Scientific integrity has widespread support
- In Australia there are 2 funding agencies and research data must be accessible; U New South Wales; Fedora is critical to this effort
  - What we pay for software to access collections is high--has 7 people working on Fedora; Tyler's kind of arguments can help illustrate to people why this is important.

Wednesday, March 13, afternoon sessions

Attendees were asked to have table discussions and report out their answers to the following questions:

1. What did you hear today?
2. What do you have concerns about?
3. What has inspired you?
4. What unanswered questions do you have?

Table reports

Fedora Futures session

Eddie via Skype from Singapore: "The sun never sets on the Fedora empire."

Jonathan: Why are we talking about FF now?

- Interesting model for doing this kind of work
- Born at OR2012
- At the end of the conference "community-managed projects" concept was discussed
- Towards the end of the conf about 8 stakeholders came to DuraSpace and said we need to talk about Fedora improvements now
- Functionally and technically it could not meet emerging use cases
- Very old code and hard to get new developers to work on Fedora
- Financial and developer resources were pledged by this group
- Still in early phases of the FF project
- Not happening in a dark room; happening among all community; trying to find ways to bring in the larger community
- Current Fedora 3 will morph into Fedora 4
- Will be talking about governance tomorrow about Fedora

Tom Cramer, Project overview

- Funding history of Fedora
- 2000-2008 with 2.4M from the Mellon Foundation; $500,000 from UVA Library
- Now work with distributed committers from 10 institutions
- Fedora is a success; arch is flexible and extensible
- Provides support for durability
- One foot in the linked data world
- decade of maturity and proven use
- Sub-community of adopters contrib and vendors
- More than a decade of use
- Widespread enterprise repositories
- Rainbows and ponies all around
- Looming storm clouds
- Nervousness on part of big universities
- Never been easy to use
- Code base is old
- Developer contributions have fallen off
- Hard to recruit new developers
- Not a sexy thing because it's mature
- Beginning to realize there is not a crisis now, but there will be in 2 years
- Other web services DB are emerging
- Chinese character for crisis and danger is also opportunity
- New front-ends
- New web arch and horizontal scaling
- Data management mandates are a good fit
- Fedora 4 "preservation enabled"; integration with AWS Glacier
- Challenges
- Communications--how to talk about it in the community; engaging people at all levels in non-technical terms
- Meta-issues; digital preservation issues in general
- How is money flowing?
- All money we are collecting for FF goes to Fedora development; last year was paid for by sponsorship; sub-project under Fedora
- Is Fedora Futures the right name? Discussion about not fragmenting the community--interest in making sure there is a path forward
- FF group would like to see the current governance model subsumed by what comes out of this meeting
- FF is the steering and vision around what Fedora should be; vision for DSpace has not been articulated
- Challenge is that DuraSpace will not do that for you; governance could be the answer or the way forward to create a vision
- Cost of ownership piece--too important to fail; curious about what proportion of F3 is in F4; If the 5.4M was not enough to endow the project what is enough
- If we know what the community needs we can scope it out then we can plan for it; what do we want fedora to be over the next 10 years?

Eddie Shin, Tech overview

- Existing set of things that we have found out about from the community that we are working on
- Put 95% of F3 in 2 days into F4
- Not positive that we should carry some things forward lie XCML in its current form
- Why not do more roadmap development; trying to avoid the mistakes of the past of going dark and then doing the big reveal to find out that it's not what you want
- We should build quickly with quantifiable experiments and get early feedback from the community (Eddie)
- Metrics and reporting: as academia changes how people are evaluated then we get to be part of the infrastructure of the university; are you working with others on what a usage report would look like? Should be part of requirements development
- Partial effort to gather input from users--survey or something?
- Phone, wiki, conf call are not that good?
- Should we do a face-to-face hackathon for non-developers; need a deep conversation about what the user community needs

Thursday, March 14, morning sessions

DSpace Session

Initial Discussion Questions:

- **What does the project need to be successful?**
  - **financial contributions**
  - **developer contributions**
  - **articulation of requirements**
- **How to increase engagement with the community?**
  - **new models of funding (membership?)**
  - **stronger project governance?**
  - **other forms of communication?**

Greatest needs in the community summary:

1) Need a roadmap - need a vision of what DSpace should be in the next 3 to 5 years

- DSpace Committers are busy with bug fixes, releases and reviewing the many contributions – making sure they play well together - don't really have time to focus on the evolution of DSpace architecture and long term road map

2) Need governance for stakeholders to go along with additional funding and long term roadmap

DSpace Futures

- discussions came down to three main projects - hope was that DSpace futures would turn into something like Fedora Futures, instead seemed to have turned into other smaller functionality projects, not necessarily a roadmap
  - REST API
  - DSpace + Hydra
  - Metadata Improvements
  - not very many volunteers so far
- Fedora community may have more developer resources, many DSpace institutions may not have developer resources

How do we energize the DSpace community?

- Is DSpace a product or project?
- more of a push to develop a DSpace user community more at conferences - set up BoF at your next conference
- more participation in DCAT and ambassador group
- need specifics to ask for money
- need community to be in charge of what changes need to be made, actionable items and identify people charged with doing it
- have a membershiop model for smaller institutions to fund DSpace improvements
  - maybe different funding model, e.g.,
    - tie requisition to deliverable
    - adoption, patch, upgrade
    - "bill me" size of institution
    - membership depending on size of instutition
- need community to specify what needs to be done - need specifics to ask for development money, tying a specific deliverable to $$
- everyone looking at DuraSpace to create roadmap, but the community needs to lead
- DSpace community

Should DSpace be UI that sits on top of Fedora?

- Fedora 4 will still not be an out-of-the box system, no UI
- should there be a DSpace Hydra head or some other type of Fedora/DSpace integration?
  - DSpace-like Hydra head: scholarSphere (PSU) and hydris (Stanford)
- is it a grant project to merge DSpace with Fedora?
- DuraSpace should scope the project

DSpace community is bifurcated community

- smaller institutions with limited resources - folks without IT support - could hosted DSpaceDirect address this need (SAAS)
- other community
  - institutions with needs where they've outgrown DSpace platform
  - contemplating exit strategy
  - need to contribute to re-envision / modify DSpace

Governance
- need a steering group
- could DCAT could serve in product-owner role - representing non-sponsors?
- centralized technical vision and technical leadership
  - how to ensure that features integrate well?
  - do we need a project director role - either funded or contributed in-kind by the community

- Roadmap
  - Scoping project to see if DSpace-on-Fedora is viable
    - DuraSpace should scope (David)
    - Need "dead-simple" migration path

- David Lewis
  - Hosting meeting at ACRL
    - MK will be there with project proposal

- will pass the hat

- Need repository managers involved in strategic vision
- steering group needs mix of technologists, administrators, repository managers

not an enterprise system, doesn't scale

VIVO Breakout Session Mike Conlon, recorder

Attendees: Mike Conlon (UF), Bill Barnett (IU), Hal Warren (APA), Julia Trimmer (Duke), Mike Bolton (Texas A&M), Dean Krafft (Cornell), Robert McDonald (Indiana), Mike Winkler (Penn), Daniel Calto (Elsevier), Jonathan Breeze (Symplectic), Alex Viggio (Digital Science), Paul Albert (Weill Cornell), Delphine Canno (Temple, by Skype), Andrew Ashton (Brown), Jon Corson-Rikert (Cornell, by video)

Three discussion areas:

1. VIVO Roadmap – Dean Krafft, facilitator
2. VIVO Strategy – Jonathan Markow, Facilitator
3. VIVO Strategy – Bill Barnett, Facilitator

VIVO Roadmap -- Dean Krafft, Facilitator

Vivo 1.6 -- VIVO-ISF, CTSAConnect, Eagle-I, bidirectional API via SPARQL, HTTP caching, search indexing, developer tools, landing page improvements. 1.6.1 running through tests.

If a grant or other sponsor can pay for a specific feature that does not take away from the strategy or roadmap, and gifts back the feature to the vivo core, there is general support.

A collection of architectural use cases -- VIVO as an engine, an aggregator, an information representation, as data store used for external apps, Balance use cases in the development road map.

Need for documentation. Using, installing, adding code. To build development community -- Internal code needs documentation. Architecture documents. Development process. On boarding documentation.

Some possible new features:

- Control the rank ordering in search results
- Linked data for libraries will need billion triple scale stores and support for commercial triple stores.
- Internationization. Labeling. Or in operation of the software.
- Archival vivo.
- External URIs.
- Ontology editor improvements, application ontology, ISF module support
- Multi-site search
- Vivo linkages

There will be an annual survey at the Implementation Fest.

How is the roadmap developed? Users, sponsors, technical. Interaction with grant funded effort. Community of developers. Balancing allocation. Sponsors, steering. Ideas -> proposals -> vetting -> deciding

Time-based vs feature-based release? Most prefer time-based release. Easier to plan. Possibly bi-annual.

Action items

Project Director hire. Community assessment of desirable features before Austin? Time at Austin to disucss? Roadmap proposal. Goals: Functioning organization, roadmap development

Governance -- Jonathan Markow, Facilitator

Reviewed the project charter describing governance and membership.

Spending incubation period developing governance and membership. Founding members are Platinum members with designation as Founders. Other levels range from bronze at $2500 to Diamond at $30,00. Membership campaign in the spring -- we have a handful of non-founders currently.

Recognize in-kind contributions. 1/2 developer would be same as platinum, on the leadership group.

Steering committee -- Bill, Paul, Dean, Jon, Mike, Kristi, Robert, Jonathan, Project Director. Should be elected by the leadership group.

Leadership group -- platinum and above. Elect the steering group. Clarify the role of the leadership group regarding strategy, budget. Intended to function as a board.

Project members -- submit requests, vote for two at large leadership group members

Work groups – largely volunteer effort with focus on an aspect of VIVO – implementation, core development, apps & tools, ontology, outreach

Duraspace – shepard the development of all things VIVO – software, community, process

Use cases for governance. How do we create a roadmap? How do we develop a strategic plan? How do we develop and approve a budget? Allocate resources? How (when/process) do we name the steering committee?

Input rights -- bronze and silver

Decision rights -- platinum and diamond

Project members at the bronze level. Consortia membership is possible to encourage participation outside the US. Learn from other projects and Duraspace.

VIVO Strategy -- Bill Barnett, Facilitator

VIVO business model -- membership, committed people, project director, Duraspace, alignment with institutuonal interests.

Kernel, application, ontology, applications

Community

What is our business? competitors? Public, shareable, Machine consumable, human attribute storage. With lineage, trust. Create and Share data about scholarly work.

Alliances. People who create and share scholarly work. We're not in the publishing business, nor administrative function, nor repository businesses.

The concept of VIVO data and promoting the production of VIVO data by any means.

Spurring adoption --early adopters, secondary messaging, early middle (value propositions), late middle (everyone else is doing it, compliance). Enabling implementation, changing the ROI. Having usage and implementation stories at various scales.