Building the iRODS Consortium

One project's journey from academic code to enterprise software

Presented at *All Things Open 2014*
October 22, 2014

Dan Bedard (danb@renci.org)
The iRODS Consortium

What's this all about?

What's this all about?

- iRODS software manages around 100 PB of data worldwide.
- (It's open source.)
- It started as a sponsored research project at UCSD.
- Now it's managed by the iRODS Consortium.

This talk is about...

This talk is about...

- What iRODS is.
- How and why the iRODS Consortium came to be.
- How the Consortium works.
- Lessons learned and challenges.

This talk is about...

- What iRODS is.
- How and why the iRODS Consortium came to be.
- How the Consortium works.
- Lessons learned and challenges.

- Thirty minutes long.

But first...

But first...

- Dan Bedard is not a software developer.

But first...

- Dan Bedard is not a software developer.
- The iRODS Consortium is based at RENCI.
- RENCI is based at the University of North Carolina at Chapel Hill.
- RENCI is an applied research organization that helps other departments at UNC.

What is iRODS?

iRODS is open source middleware for...

Data Discovery,

Workflow Automation,

Secure Collaboration,

and Data Virtualization.

What is iRODS?

iRODS is open source middleware for...
↑sits between the user/admin and the file system

Data Discovery, ←metadata annotation

Workflow Automation, ←über cron

Secure Collaboration, ←consolidates access and control of data across sites

and Data Virtualization. ←all your storage in a single namespace

What is iRODS?

iRODS makes huge sets of unstructured data manageable, usable, and shareable.

What is iRODS?

iRODS makes huge sets of unstructured data manageable, usable, and shareable.

Broad usage in…

- genomics and life sciences (Sanger, Broad, BGI, Lineberger)
- large scientific data sets (NASA, NOAA, NAOA)
- digital libraries (French National Library, SNIC)

What is iRODS?

iRODS makes huge sets of unstructured data manageable, usable, and shareable.

Broad usage in…

- genomics and life sciences (Sanger, Broad, BGI, Lineberger)
- large scientific data sets (NASA, NOAA, NAOA)
- digital libraries (French National Library, SNIC)

irods.org

iRODS History, Early Years

iRODS History, Early Years

- Storage Resource Broker (SRB) developed by General Atomics, Data Intensive Cyber Environments group (DICE) at UCSD, and San Diego Supercomputer Center (SDSC) under DARPA funding.

- Forked into:
  - a closed source commercial product and
  - a free non-commercial product. (Source code was available upon request.)

iRODS History, Early Years

- Storage Resource Broker (SRB) developed by General Atomics, Data Intensive Cyber Environments group (DICE) at UCSD, and San Diego Supercomputer Center (SDSC) under DARPA funding.

- Forked into:
  - a closed source commercial product and
  - a free non-commercial product. (Source code was available upon request.)

- DICE deprecates SRB. Builds a new system around a *"rule engine"* and re-writes SRB concepts into a new system. Calls it iRODS.

- The iRODS rule engine enables policy definition. (Any condition, any action.)

iRODS History, Early Years

- Storage Resource Broker (SRB) developed by General Atomics, Data Intensive Cyber Environments group (DICE) at UCSD, and San Diego Supercomputer Center (SDSC) under DARPA funding.

- Forked into:
  - a closed source commercial product and
  - a free non-commercial product. (Source code was available upon request.)

- DICE deprecates SRB. Builds a new system around a *"rule engine"* and re-writes SRB concepts into a new system. Calls it iRODS.

- The iRODS rule engine enables policy definition. (Any condition, any action.)

- DICE group expands. Some staff at UNC, some at UCSD.

iRODS History, Early Years

- Storage Resource Broker (SRB) developed by General Atomics, Data Intensive Cyber Environments group (DICE) at UCSD, and San Diego Supercomputer Center (SDSC) under DARPA funding.

- Forked into:
  - a closed source commercial product and
  - a free non-commercial product. (Source code was available upon request.)

- DICE deprecates SRB. Builds a new system around a *"rule engine"* and re-writes SRB concepts into a new system. Calls it iRODS.

- The iRODS rule engine enables policy definition. (Any condition, any action.)

- DICE group expands. Some staff at UNC, some at UCSD.

- An inflection point.

iRODS in 2011

iRODS in 2011

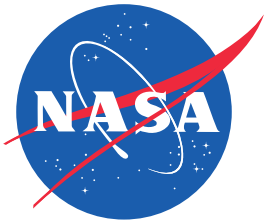
EarthCube Layered Architecture	NSF	4/1/2012 – 3/31/2013
DFC Supplement for Extensible Hardware	NSF	9/1/2011 – 8/31/2015
DFC Supplement for Interoperability	NSF	9/1/2011 – 8/31/2015
DataNet Federation Consortium	NSF	9/1/2011 – 8/31/2016
SDCI Data Improvement	NSF	10/1/2010 – 9/30/2013
Subcontract: Temporal Dynamics of Learning Center	NSF	1/1/2010 – 12/31/2010
National Climatic Data Center	NOAA	10/1/2009 – 9/1/2010
NARA Transcontinental Persistent Archive Prototype	NSF	9/15/2009 – 9/30/2010
Subcontract: Temporal Dynamics of Learning Center	NSF	3/1/2009 – 12/31/2009
Transcontinental Persistent Archive Prototype	NSF	9/15/2008 – 8/31/2013
Petascale Cyberfacility for Seismic Community	NSF	4/1/2008 – 3/30/2010
Data Grids for Community Driven Applications	NSF	10/1/2007 – 9/30/2010
Joint Virtual Network Centric Warfare	DOD	11/1/2006 – 10/30/2007
Petascale Cyberfacility for Seismic	NSF	10/1/2006 – 9/30/2009
LLNL Scientific Data Management	LLNL	3/1/2005 – 12/31/2008
NARA Persistent Archives	NSF	10/1/2004 – 6/30/2008
Constraint-based Knowledge	NSF	10/1/2004 – 9/30/2006
NDIIPP California Digital Library	LC	2/1/2004 – 1/31/2007
NASA Information Power Grid	NASA	10/1/2003 – 9/30/2004
National Science Digital Library	NSF	10/1/2002 – 9/30/2006
NARA Persistent Archive	NSF	6/1/2002 – 5/31/2005
SCEC Community Modeling	NSF	10/1/2001 – 9/30/2006
Particle Physics Data Grid	DOE	8/15/2001 – 8/14/2004
Grid Physics Network	NSF	7/1/2000 – 6/30/2005
Digital Library Initiative UCSB	NSF	9/1/1999 – 8/31/2004
Digital Library Initiative Stanford	NSF	9/1/1999 – 8/31/2004
Persistent Archive	NARA	9/1999 – 8/2000
Information Power Grid	NASA	10/1/1998 – 9/30/1999
Terascale Visualization	DOE	9/1/1998 – 8/31/2002
Persistent Archive	NARA	9/1998 – 8/1999
NPACI data management	NSF	10/1/1997 – 9/30/1999
DOE ASCI	DOE	10/1/1997 – 9/30/1999
Distributed Object Computation Testbed	DARPA/USPTO	8/1/1996 – 12/31/1999
Massive Data Analysis Systems	DARPA	9/1/1995 – 8/31/1996

iRODS in 2011

renci

RESEARCH \ ENGAGEMENT \ INNOVATION

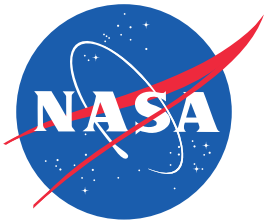
iRODS in 2011



We want to manage (even more) massive amounts of climate data using iRODS…

but we have to do a security audit for each new version.

iRODS in 2011



We want to manage (even more) massive amounts of climate data using iRODS...

but we have to do a security audit for each new version.

We're thinking about funding iRODS-based federation between research groups...

What's your plan for long-term sustainability?

iRODS in 2011

iRODS was used in many research groups worldwide because:

- It does things that no other software does…

iRODS in 2011

iRODS was used in many research groups worldwide because:

- It does things that no other software does…
- for free.

iRODS in 2011

iRODS was used in many research groups worldwide because:

- It does things that no other software does…
- for free.
- And the DICE group was able to support the needs of the user community.

iRODS in 2011

iRODS was used in many research groups worldwide because:

- It does things that no other software does…
- for free.
- And the DICE group was able to support the needs of the user community.

But…

iRODS in 2011

iRODS was used in many research groups worldwide because:

- It does things that no other software does…
- for free.
- And the DICE group was able to support the needs of the user community.

But…

- Today's grant funding doesn't provide for tomorrow's support.

iRODS in 2011

iRODS was used in many research groups worldwide because:

- It does things that no other software does…
- for free.
- And the DICE group was able to support the needs of the user community.

But…

- Today's grant funding doesn't provide for tomorrow's support.
- The iRODS architecture could be unwieldy.

iRODS in 2011

iRODS was used in many research groups worldwide because:

- It does things that no other software does…
- for free.
- And the DICE group was able to support the needs of the user community.

But…

- Today's grant funding doesn't provide for tomorrow's support.
- The iRODS architecture could be unwieldy.
- And the development and support community might be difficult to scale.

iRODS in 2011

Stan Ahalt
Director@RENCI

iRODS is an asset to RENCI.

iRODS in 2011

Stan Ahalt
Director@RENCI

iRODS is an asset to RENCI.

And lots of people are using it.

iRODS in 2011

Stan Ahalt
Director@RENCI

iRODS is an asset to RENCI.

And lots of people are using it.

But no one will pay to sustain it.

iRODS in 2011

Stan Ahalt
Director@RENCI

- iRODS is an asset to RENCI.
- And lots of people are using it.
- But no one will pay to sustain it.
- Unless…

iRODS in 2011

Stan Ahalt
Director@RENCI

We need an enterprise-ready iRODS.

iRODS History: 2012-2014

- RENCI forks and creates E-iRODS. Forms the iRODS Consortium.

.
.
.

- E-iRODS and Community iRODS merged into iRODS 4.0.

iRODS History: 2012-2014

- RENCI forks and creates E-iRODS. Forms the iRODS Consortium.

.
.
.

- E-iRODS and Community iRODS merged into iRODS 4.0.

iRODS 4.0 and beyond are enterprise-ready.

- Modern software development practices.
- Pluggable architecture.
- Packaged installation.
- More extensive testing and continuous integration.

The iRODS Consortium

- Founded after discussion with a Major Stakeholder.
- Founding members would be RENCI, DICE, and that Major Stakeholder.

The iRODS Consortium

- Founded after discussion with a Major Stakeholder.
- Founding members would be RENCI, DICE, and that Major Stakeholder.

- Operating model based on the Kerberos Foundation.
  - Tests and certifies outside vendors' code.

The iRODS Consortium: Funding Model

- iRODS stakeholders join to protect their infrastructure investment.
  - Four levels of membership ($10k to $150k annually).
  - Increasing influence (voting) and priority (support).

The iRODS Consortium: Funding Model

- iRODS stakeholders join to protect their infrastructure investment.
  - Four levels of membership ($10k to $150k annually).
  - Increasing influence (voting) and priority (support).

- Additionally: system integration, standby support, and training.

The iRODS Consortium: Governance

- - Approves budgets, staffing, major release plans
- - Develops software release roadmaps, establishes working groups
- - Discusses technical approaches, open issues

So what happened in between?

So what happened in between?

iRODS History: 2012-2014

- RENCI forks and creates E-iRODS. Forms the iRODS Consortium.

.
.
.

- E-iRODS and Community iRODS merged into iRODS 4.0.

So what happened in between?

- Creation of the Consortium Plan.

So what happened in between?

- Creation of the Consortium Plan.
- Major Stakeholder required merge before they would sign on to the Consortium.

So what happened in between?

- Creation of the Consortium Plan.
- Major Stakeholder required merge before they would sign on to the Consortium.
- March 2013: Merge plan announced.

So what happened in between?

- Creation of the Consortium Plan.
- Major Stakeholder required merge before they would sign on to the Consortium.
- March 2013: Merge plan announced.
- September 2013: Brand Fortner joins as Executive Director

So what happened in between?

- Creation of the Consortium Plan.
- Major Stakeholder required merge before they would sign on to the Consortium.
- March 2013: Merge plan announced.
- September 2013: Brand Fortner joins as Executive Director
- March 2014: iRODS 4.0 released. First Consortium members join.

So what happened in between?

- Creation of the Consortium Plan.
- Major Stakeholder required merge before they would sign on to the Consortium.
- March 2013: Merge plan announced.
- September 2013: Brand Fortner joins as Executive Director
- March 2014: iRODS 4.0 released. First Consortium members join.
- October 2014: The Consortium has six members: RENCI, DICE, DDN, Seagate, Wellcome Trust Sanger Institute, and EMC

Why not become a private company?

Partly about keeping our user community happy.

Why not become a private company?

Partly about keeping our user community happy.

Also…

$$v_{\text{escape}} = \sqrt{\frac{2GM}{R}}$$

The escape velocity for spinning out of UNC seemed too high at the time.

Why not become a private company?

Partly about keeping our user community happy.

Also…

$$v_{\text{escape}} = \sqrt{\frac{2GM}{R}}$$

The escape velocity for spinning out of UNC seemed too high at the time.

- A *lot* of work and risk involved.

Why not become a private company?

Partly about keeping our user community happy.

Also…

$$v_{\text{escape}} = \sqrt{\frac{2GM}{R}}$$

The escape velocity for spinning out of UNC seemed too high at the time.

- A *lot* of work and risk involved.
- High value of being at RENCI.

Why not become a private company?

Partly about keeping our user community happy.

Also…

$$v_{\text{escape}} = \sqrt{\frac{2GM}{R}}$$

The escape velocity for spinning out of UNC seemed too high at the time.

- A *lot* of work and risk involved.
- High value of being at RENCI.
- Longevity afforded by association with a University.

Lessons Learned

Lessons Learned

- Don't underestimate how long this will take.
  - Staying with the University was slow, but low risk.
  - Convincing the community that this is real takes time.

Lessons Learned

- Don't underestimate how long this will take.
  - Staying with the University was slow, but low risk.
  - Convincing the community that this is real takes time.

- It's difficult to change what you're doing and how you're doing it at the same time.

Lessons Learned

- Don't underestimate how long this will take.
  - Staying with the University was slow, but low risk.
  - Convincing the community that this is real takes time.
- It's difficult to change what you're doing and how you're doing it at the same time.
- Think clearly and thoroughly about what your objectives are.

Lessons Learned

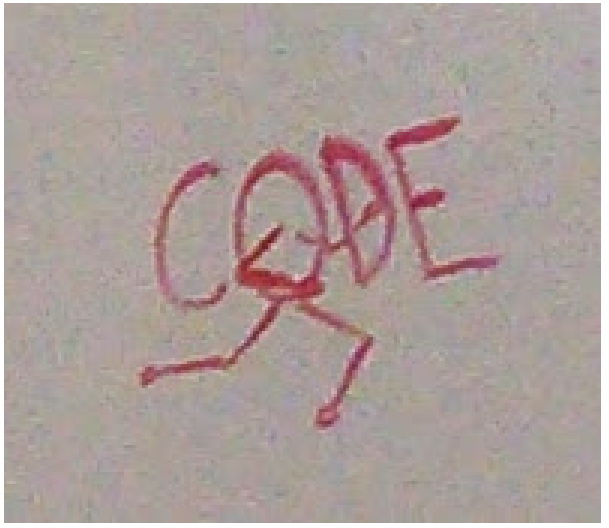
- Don't underestimate how long this will take.
  - Staying with the University was slow, but low risk.
  - Convincing the community that this is real takes time.
- It's difficult to change what you're doing and how you're doing it at the same time.
- Think clearly and thoroughly about what your objectives are.
- Structure matters.

Lessons Learned

- Don't underestimate how long this will take.
  - Staying with the University was slow, but low risk.
  - Convincing the community that this is real takes time.
- It's difficult to change what you're doing and how you're doing it at the same time.
- Think clearly and thoroughly about what your objectives are.
- Structure matters.
- You need Major Stakeholders and evangelists.

Lessons Learned

- At the end of the day, you need something that solves the problem better or cheaper than anyone else.



"Running code wins the day."

Building a Consortium

articulating the idea and vision

getting early members, establishing the community

establishing programs, working groups, bylaws, begin to show value & build momentum

new programs, growing existing programs, increasingly member-driven as founders give up some control

A sustainable, member-driven organization, possibly with professional managers/administrators

Building a Consortium

articulating the idea and vision

getting early members, establishing the community

establishing programs, working groups, bylaws, begin to show value & build momentum

We think we're right around here.

new programs, growing existing programs, increasingly member-driven as founders give up some control

A sustainable, member-driven organization, possibly with professional managers/administrators

"Last Slide Stuff"

- Thank you to:
  - David Knowles, Charles Schmitt, Stan Ahalt, Jason Coposky, and Terrell Russell for historical context.
  - The entire iRODS Consortium team for their continuing efforts.

- These slides licensed under Creative Commons

http://creativecommons.org/licenses/by-sa/4.0/

- These slides made with http://remarkjs.com/

Questions?

Suggestions?