VisTrails: Using Provenance and Workflows for Scientific Exploration

David Koop
School of Engineering, New York University
Orchestrate Different Tools & Libraries
Orchestrate Different Tools & Libraries

OpenMP  OpenCL

\[ y_{it} = \beta x_{it} + \mu_i + \epsilon_{it} \]
Abstract & Structure Computations

[Paul Klee]
Structured & Abstracted Code

data = vtk.vtkStructuredPointsReader()
data.SetFileName(../examples/data/head.120.vtk)

contour = vtk.vtkContourFilter()
contour.SetInput(data.GetOutput())
contour.SetValue(0, 67)

mapper = vtk.vtkPolyDataMapper()
mapper.SetInput(contour.GetOutput())
mapper.ScalarVisibilityOff()

actor = vtk.vtkActor()
actor.SetMapper(mapper)

cam = vtk.vtkCamera()
cam.SetViewUp(0, 0, -1)
cam.SetPosition(745, -453, 369)
cam.SetFocalPoint(135, 135, 150)
cam.ComputeViewPlaneNormal()

ren = vtk.vtkRenderer()
ren.AddActor(actor)
ren.SetActiveCamera(cam)
ren.ResetCamera()
renwin = vtk.vtkRenderWindow()
renwin.AddRenderer(ren)

style = vtk.vtkInteractorStyleTrackballCamera()
iren = vtk.vtkRenderWindowInteractor()
iren.SetRenderWindow(renwin)
iren.SetInteractorStyle(style)
iren.Initialize()
iren.Start()
Ensure Quality, Reliability, & Trustworthiness

[Pentium Chip with FDIV Bug, Photo by Konstantin Lanzet, CC BY-SA 3.0]
Provenance in Art

Rembrandt van Rijn
Dutch, 1606 - 1669

Self-Portrait, 1659
oil on canvas
Andrew W. Mellon Collection
1937.1.72

Provenance


[1] This early provenance is established by presence of a mezzotint after the portrait by R. Earlom (1743-1822), dated 1767. See John Charrington, A Catalogue of the Mezzotints After, or Said to Be After, Rembrandt, Cambridge, 1923, no. 49.

Associated Names

• Buccleuch, Henry, 3rd Duke of
• Buccleuch, John Charles, 7th Duke of
• Colnaghi & Co., Ltd., P. & D.
• Knoedler & Company, M.
• Mellon, Andrew W.
• Mellon Educational and Charitable Trust, The A.W.
• Montagu, and 4th Earl of Cardigan, George, 3rd Duke of
Provenance in Art

Rembrandt van Rijn
Dutch, 1606 - 1669
Self-Portrait, 1659
oil on canvas
Andrew W. Mellon Collection
1937.1.72

Provenance

[1] This early provenance is established by presence of a mezzotint after the portrait by R. Earlom (1743-1822), dated 1767. See John Charrington, A Catalogue of the Mezzotints After, or Said to Be After, Rembrandt, Cambridge, 1923, no. 49.

Associated Names
- Buccleuch, Henry, 3rd Duke of
- Buccleuch, John Charles, 7th Duke of
- Colnaghi & Co., Ltd., P. & D.
- Knoedler & Company, M.
- Mellon, Andrew W.
- Mellon Educational and Charitable Trust, The A.W.
- Montagu, and 4th Earl of Cardigan, George, 3rd Duke of
Provenance in Science

• Provenance: the **lineage** of data, a computation, or a visualization

• **Provenance is as (or more) important as the result!**

• Old solution:
  - Lab notebooks

• New problems:
  - Large volumes of data
  - Complex analyses
  - Writing notes doesn’t scale

[DNA Recombination, Lederberg]
Provenance in Science

• Provenance: the **lineage** of data, a computation, or a visualization

• **Provenance is as (or more) important as the result!**

• Old solution:
  - Lab notebooks

• New problems:
  - Large volumes of data
  - Complex analyses
  - Writing notes doesn’t scale

[DNA Recombination, Lederberg]
Reproducibility
Reproducibility

- Capture **how** results were achieved
- Includes **many** different items
- Improve **collaboration** and sharing
Reuse Past Work

[PETCO Park, Photo by Edward O’Connor, CC BY-SA 2.0]
Reuse Past Work

Database of Workflows

Thursday, August 14, 14
Unencumbered Exploration & Learning
Data Exploration

[Modified from Van Wijk, Vis 2005]
Data Exploration

• Data analysis and visualization are iterative processes
• In exploratory tasks, change is the norm!

[Modified from Van Wijk, Vis 2005]
Exploration and Creativity Support

• Reasoning is key to the exploratory processes

• “Reflective reasoning requires the ability to store temporary results, to make inferences from stored knowledge, and to follow chains of reasoning backward and forward, sometimes backtracking when a promising line of thought proves to be unfruitful. ...the process is slow and laborious” — Donald A. Norman

• Need external aids—tools to facilitate this process
  - Creativity support tools [Ben Shneiderman]

• Need aid from people—collaboration
VisTrails
VisTrails

- Comprehensive **provenance infrastructure** for computational tasks
- Focus on **exploratory** tasks such as simulation, visualization, and data analysis
- **Transparency**y tracks provenance of the discovery process—from data acquisition to visualization
  - The **trail** followed as users generate and test hypotheses
  - Users can refer back to any point along this trail at any time
- Leverage provenance to **streamline exploration**
- Focus on **usability**—build tools for scientists
VisTrails

- Open-source, freely downloadable system (www.vistrails.org)
  - Also on GitHub (github.com/vistrails) and PyPI
- Multi-platform: users on Mac, Linux, and Windows
- Python code and uses PyQt and Qt for the interface
- Over 35,000 downloads
- User’s guide, wiki, and mailing list
- Many users in different disciplines and countries:
  - Visualizing environmental simulations (CMOP STC)
  - Simulation for solid, fluid and structural mechanics (Galileo Network, UFRJ Brazil)
  - Quantum physics simulations (ALPS, ETH Zurich)
  - Climate analysis (UV-CDAT, LLNL)
  - Habitat modeling (USGS)
  - Open Wildland Fire Modeling (U. Colorado, NCAR)
  - High-energy physics (LEPP, Cornell)
  - Cosmology simulations (LANL)
  - Using tms for improving memory (Psychiatry, U. Utah)
  - eBird (Cornell, NSF DataONE)
  - Astrophysical Systems (LSU)
  - NIH NBCR (UCSD)
  - Pervasive Technology Labs (Indiana University)
  - Linköping University
  - University of North Carolina, Chapel Hill
  - UTEP
SAHM: Modeling the Spread of Invasive Species

b_{rt_1} Probability

Glm_1 Probability

Mars_1 Probability

Rf_1 Probability

ROC Plots for Cross-Validation
SAHM: Modeling the Spread of Invasive Species

[J. Morisette et al., USGS-Fort Collins, NASA]
UV-CDAT: Climate Science
UV-CDAT: Climate Science

- CDMS_FileReader
- CDMSVariable
- CDMS_VolumeReader
- VolumeRenderer
- MapCell3D
- genutil.averager (CDMSUnaryVariableOperation)
- Calculator (CDMSUnaryVariableOperation)
- CellLocation
- CDMSYxvsx
- CDMSCell
UV-CDAT: Climate Science

[D. N. Williams, T. Maxwell, E. Santos, et al., LLNL, NASA, NYU]
Example: MTA Fare Data

Parameters

- HTTPFile.url: web.mta.info/.../fares_130824.csv
- CSVFile.skip_lines: 2
- JoinTables.left_col: STATION
- JoinTables.right_col: _key
- MPLAxesProps.xlabel: Full Fares Purchased

David Koop
ATPESC 2014

Thursday, August 14, 14
Example: MTA Fare Data

Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTPFile.url</td>
<td>web.mta.info/.../fares_130824.csv</td>
</tr>
<tr>
<td>CSVFile.skip_lines</td>
<td>2</td>
</tr>
<tr>
<td>JoinTables.left_col</td>
<td>STATION</td>
</tr>
<tr>
<td>JoinTables.right_col</td>
<td>_key</td>
</tr>
<tr>
<td>MplAxesProps.xlabel</td>
<td>Full Fares Purchased</td>
</tr>
</tbody>
</table>
Example: MTA Fare Data

Parameters

- HTTPFile.url: web.mta.info/.../fares_130824.csv
- CSVFile.skip_lines: 2
- JoinTables.left_col: STATION
- JoinTables.right_col: _key
- MplAxesProps.xlabel: Full Fares Purchased

David Koop

ATPESC 2014

Thursday, August 14, 14
Example: MTA Fare Data

HTTPFile

CSVFile

TableCell

JoinTables

GetFareData

DateRange

Parameters

HTTPFile.url

web.mta.info/.../fares_130824.csv

CSVFile.skip_lines

2

JoinTables.left_col

STATION

JoinTables.right_col

_key

MplAxesProps.xlabel

Full Fares Purchased

ProjectTable

GMapCell

MplFigureCell

MplAxesProperties

MplBar

MplFigure

BuildLabels

(PythonSource)

Map

DateRange

(PythonSource)
Example: MTA Fare Data

Parameters
- **HTTPFile.url**: web.mta.info/.../fares_130824.csv
- **CSVFile.skip_lines**: 2
- **JoinTables.left_col**: STATION
- **JoinTables.right_col**: _key
- **MplAxesProps.xlabel**: Full Fares Purchased

**Data Sources**
- **HTTPFile**
- **CSVFile**
- **JSONFile**

**Join Tables**
- **JoinTables**

**Visualization**
- **MplBar**
- **MplFigure**
- **MplAxesProperties**
- **MplBar**

**Other**
- **GMapCell**

David Koop
ATPESC 2014
NYU Polytechnic School of Engineering
Thursday, August 14, 14
Example: MTA Fare Data

Parameters

- **HTTPFile.url**: web.mta.info/.../fares_130824.csv
- **CSVFile.skip_lines**: 2
- **JoinTables.left_col**: STATION
- **JoinTables.right_col**: _key
- **MplAxesProps.xlabel**: Full Fares Purchased

David Koop

Thursday, August 14, 14
Capturing Provenance
Workflow Evolution Provenance of MTA Fare Data

- Initial data
  - Corrected data
    - Station locations
      - Station map
      - Added fares
      - Difference
  - NOV 16 data
  - August 16 Tab
  - Full fares map
  - Broadway line
  - August 16
    - Broadway diff map
  - Sum of ffs
    - 30-D weekly
      - 161st-River
        - 30-D weekly
          - Sum of ffs

- GetFareData (PythonSource)
- DateRange (PythonSource)
- BuildLabels (PythonSource)
- MplBar
- MplAxesProperties
- MplFigure
- MplFigureCell
- MplFigureProperties
- GMapCell
- GMapCell
Workflow Evolution Provenance of MTA Fare Data

- Initial data
- Station locations
- Corrected data
- Station map
- Added fares
- November ff
- November 2 data
- August 16 Tab
- Full fares map
- Difference
- Broadway line
- August 16
- Broadway diff map
- Sum of ffs
- 30-D weekly
- 161st-River
- With labels
- Concourse line
- Filtered
- Heatmap

David Koop
ATPESC 2014

Thursday, August 14, 14
Workflow Evolution Provenance of MTA Fare Data
Workflow Evolution Provenance

...
delete module “GMapCell”
delete module “CellLocation”
delete module “ProjectTable”
delete module “SelectFromTable”
...
add module “SelectFromTable”
add parameter “float_expr” to “SelectFromTable”
    with value “latitude > 40.6”
delete parameter “float_expr” from “SelectFromTable”
add parameter “float_expr” to “SelectFromTable”
    with value “latitude > 40.7”
delete parameter “float_expr” from “SelectFromTable”
add parameter “float_expr” to “SelectFromTable”
    with value “latitude > 40.8”
...
Execution Provenance
Workflow Upgrades
Workflow Upgrades
Workflow Upgrades
Workflow Upgrades
Provenance of Workflow Upgrades

Change-based Provenance:

- delete connection StringToNumeric → AggregateData
- delete connection AggregateData → AggregateData
- delete connection AggregateData → JoinData
- delete connection JoinData → ExtractColumn
- delete connection JoinData → ExtractColumn
- delete connection ExtractColumn → MplScatterplot
- delete connection ExtractColumn → MplScatterplot
- delete connection MplScatterplot → MplFigure
- delete connection MplFigure → MplFigureCell
- delete module AggregateData version 1.0.4
- delete module AggregateData version 1.0.4
- delete module ExtractColumn version 0.9.7
- delete module ExtractColumn version 0.9.7
- delete module MplScatterplot version 2.0.0
- delete module MplScatterplot version 2.0.0
- delete module MplFigure version 2.0.0
- delete module MplFigure version 2.0.0
- delete module MplFigureCell version 2.0.0
- add module ComposeData version 1.1.0
- add module ExtractColumn version 1.0.2
- add module ExtractColumn version 1.0.2
- add module MplScatterplot version 2.0.1
- add module MplFigure version 2.0.1
- add module MplFigureCell version 2.0.1
- add connection StringToNumeric → ComposeData
- add connection ComposeData → JoinData
- add connection JoinData → ExtractColumn
- add connection JoinData → ExtractColumn
- add connection ExtractColumn → MplScatterplot
- add connection ExtractColumn → MplScatterplot

...
Adding Provenance to 3rd-Party Tools

Autodesk Maya
Adding Provenance to 3rd-Party Tools
Adding Provenance to 3rd-Party Tools

Autodesk Maya

ParaView

VisIt
Adding Provenance to 3rd-Party Tools

Autodesk Maya

ParaView

VisIt

ImageVis3d
VisTrails Provenance Plugin for ParaView
VisTrails Provenance Plugin for ParaView
Querying and Re-using Provenance
Querying Provenance

- What process led to the output image?
- What input datasets contributed to the output image?
- What workflows include resampling and isosurfacing with isovalue 57?
- Graph traversal or graph patterns
  - How do we write such queries?
Querying Provenance by Example

• Provenance is represented as graphs: hard to specify queries using text!
• Querying workflows by example [Scheidegger et al., TVCG 2007; Beeri et al., VLDB 2006; Beeri et al. VLDB 2007]
  - WYSIWYQ -- What You See Is What You Query
  - Interface to create workflow is same as to query
Linking Provenance and Data

- Filenames are often the mode of identification in data exploration
- We might also use URIs or access curated data stores
  - Can this always be expected for exploratory tasks?
  - What happens if offline?
- Solution:
  - Managed store for data associated with computations
  - Improved data identification
  - Automatic versioning
Linking Provenance and Data

- Filenames are often the mode of identification in data exploration
- We might also use URIs or access curated data stores
  - Can this always be expected for exploratory tasks?
  - What happens if offline?
- Solution:
  - Managed store for data associated with computations
  - Improved data identification
  - Automatic versioning

```xml
<workflow_exec id="1">
  <m_exec id="5">
    name="vtkStructuredDataReader"
    package="edu.utah.sci.vistrails.vtk"
    version="5.6.0">
    <param id="2" name="SetFile" value="/MyData/05-12-sc2.dat"/>
  </m_exec>

  <m_exec id="6">
    name="vtkContourFilter"
    package="edu.utah.sci.vistrails.vtk"
    version="5.6.0">
    <param id="3" name="SetValue" value="[1, 57]"/>
    <param id="4" name="ComputeScalarsOn" value="True"/>
  </m_exec>

  ...

  <m_exec id="11">
    name="FileSink"
    package="edu.utah.sci.vistrails.basic"
    version="1.5">
    <param id="15" name="path" value="/home/a/results/23.out"/>
  </m_exec>
</workflow_exec>
```

FILE NOT FOUND
Linking Provenance and Data

- Filenames are often the mode of identification in data exploration
- We might also use URIs or access curated data stores
  - Can this always be expected for exploratory tasks?
  - What happens if offline?
- Solution:
  - Managed store for data associated with computations
  - Improved data identification
  - Automatic versioning
Full Data Provenance

newfilename.dat

HASH CONTENTS

0ab678cd...

QUERY FILE STORE

12ab3-45ef2...

OBTAIN INPUT REFS

QUERY PROVENANCE

OBTAIN INPUT FILES

input files

[VisTrails Persistence Package]
Building Visualization Pipelines
Completions

[URL Completion, Safari]

[Code Completion, Intellisense]

[Web Search Completion, Google]
Visualization Pipeline Completions
VisComplete Overview

- **Mine** provenance collection: Identify graph fragments that co-occur in a collection of workflows (**Data-Driven**)
- Predict sets of likely workflow additions to a given partial workflow
Suggestion Interface
VisComplete Results
VisComplete Results

Diagram showing the relationship between various VTK (Visualization Toolkit) classes and filters, including:
- vtkDataSetReader
- vtkTransform
- vtkSphereSource
- vtkTransformFilter
- vtkRungeKutta4
- vtkArrowSource
- vtkStreamTracer
- vtkGlyph3D
- vtkTubeFilter
- vtkPolyDataMapper
- vtkOutlineFilter
- vtkActor
- vtkProperty
- vtkRender

Legend:
- User-Added
- Completed

Thursday, August 14, 14
Visualization by Analogy
Visualization by Analogy
Visualization by Analogy
Generating Visualizations by Analogy

A is to B as C is to D

PDB Report

- **Protein Title**: NEURAL CELL ADHESION MOLECULE, MODULE 2, NMR, 20 STRUCTURES
- **Authors**: P.H. JENSEN, V. SONORA, N.K. THOMSEN, V. BEREZIN, E. BOCK, F.M. POULSEN
- **Atom count**: C: 9560, H: 15440, N: 2580, O: 2680, S: 60
- **PDB Entry**: [PDB Entry]
Generating Visualizations by Analogy

A is to

as

C is to

D
Generating Visualizations by Analogy

- Compute **difference** $\Delta(A, B)$ from provenance
  - $D = \Delta(A, B) \circ C$ is often not a valid workflow
Generating Visualizations by Analogy

• Compute **difference** $\Delta(A,B)$ from provenance
  - $D = \Delta(A,B) \circ C$ is often not a valid workflow
• Find **map** between $A$ & $C$: $\text{map}(A,C)$
Generating Visualizations by Analogy

• Compute **difference** \(\Delta(A,B)\) from provenance
  - \(D = \Delta(A,B) \circ C\) is often not a valid workflow

• Find **map** between A & C: \(\text{map}(A,C)\)

• Compute **mapped difference** \(\Delta_{AC}(A,B) = \text{map}(A,C) \Delta(A,B)\)
  - \(D = \Delta_{AC}(A,B) \circ C\)
VisMashup

Acquire and Analyze Pipelines

Create Views (Simplify Pipelines)

Combine Views

App generation and deployment
Provenance in the Wild
Galois Conjugates of Topological Phases

M. H. Freedman,1 J. Guéhéneuc,2 M. B. Hastings,1 J. S. Trebst,1 M. Troyer,2 and Z. Wang1

1Microsoft Research, Station Q, University of California, Santa Barbara, CA 93106, USA
2Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland

(Dated: July 6, 2011)

Galois conjugation relates unitary conformal field theories (CFTs) and topological quantum field theories (TQFTs) to their non-unitary counterparts. Here we investigate Galois conjugates of quantum double models, such as the Levin-Wen model. While these Galois conjugated Hamiltonians are typically non-Hermitian, we find that their ground state wave functions still obey a generalized version of the usual code property (local operators do not act on the ground state manifold) and hence enjoy a generalized topological protection. The key question addressed in this paper is whether such non-unitary topological phases can also appear as the ground states of Hamiltonians. Specific attempts at constructing Hermitian Hamiltonians with these ground states lead to a loss of the code property and topological protection of the degenerate ground states. Beyond this we rigorously prove that no local change of basis (IV) can transform the ground states of the Galois conjugated doubled Fibonacci theory into the ground states of a topological model whose Hermitian Hamiltonian satisfies Lieb-Robinson bounds. These include all gapped local or quasi-local Hamiltonians. A similar statement holds for many other non-unitary TQFTs. One consequence is that the “Gaffnian” wave function cannot be the ground state of a gapped fractional quantum Hall state.

1. INTRODUCTION

Galois conjugation, by definition, replaces a root of a polynomial by another one with identical algebraic properties. For example, and are Galois conjugate (consider ) as are and as well as and . In physics Galois conjugation can be used to convert non-unitary conformal field theories (CFTs) to unitary ones, and vice versa. One famous example is the non-unitary Yang-Lee CFT, which is Galois conjugate to the Fibonacci CFT , the even (or integer-spin) subset of .

In statistical mechanics non-unitary conformal field theories have a venerable history. However, it has remained less clear if there exist physical situations in which non-unitary models can provide a useful description of the low energy physics of a quantum mechanical system -- after all, Galois conjugation typically destroys the Hermitian property of the Hamiltonian. Some non-Hermitian Hamiltonians, which surprisingly have totally real spectrum, have been found to arise as low energy models for a gapped 2D quantum mechanical system . In particular, there is currently a discussion on whether or not the “Gaffnian” wave function could be the ground state for a gapped fractional quantum Hall (FQH) state albeite. This model, which is also called “DFib”, is a topological quantum field theory (TQFT) whose states are string-nets on a surface labeled by either a trivial or “Fibonacci” anyon. From this starting point, we give a rigorous argument that the “Gaffnian” ground state cannot be locally conjugated to the ground state of any topological phase, within a Hermitian model satisfying Lieb-Robinson (LR) bounds (which includes but is not limited to gapped local and quasi-local Hamiltonians).

LR bounds are a technical tool for local lattice models. In relativistically invariant field theories, the speed of light is a strict upper bound to the velocity of propagation. In lattice theories, the LR bounds provide a similar upper bound for a velocity called the LR velocity, but in contrast to the relativistic case there can be some exponentially small “leakage” outside the light-cone in the lattice case. The LR bounds are a way of bounding the leakage outside the light-cone. The LR velocity is set by microscopic details of the Hamiltonian, such as the interaction strength and range. Combining the LR bounds with the spectral gap enables us to prove locality of various correlation and response functions. We will call a Hamiltonian a Lieb-Robinson model if it satisfies LR bounds.

We work primarily with a single example, but it should be clear that the concept of Galois conjugation can be widely applied to TQFTs. The essential idea is to retain the particle types and fusion rules of a unitary theory but when one comes to writing down the algebraic form of the F-matrices (also called ) symbols), the entries are now Galois conjugated. A slight complication, which is actually an asset, is that writing an F-matrix requires a gauge choice and the most convenient choice may differ before and after Galois conjugation.

Our method is not restricted to Galois conjugated DFib but and its factors and , but can be generalized to infinitely many non-unitary TQFTs, showing that they will not arise as low energy models for a gapped 2D quantum mechan-
Benefits of Provenance-Rich Publications

• Produce more **knowledge**—not just text
• Allow scientists to stand on the shoulders of giants (and their own)
• Science can move **faster**!
• **Higher-quality** publications
• Authors will be more **careful**
• Many eyes to **check** results
• Describe more of the discovery process: people only describe successes, can we **learn from mistakes**?
• Expose users to different techniques and tools: expedite their training; and potentially **reduce their time to insight**
Vision: Provenance-Rich Science

Data Management

Provenance

Visualization

Publishing

DATA

DATA

Computation

David Koop

ATPESC 2014

NYU POLYTECHNIC SCHOOL OF ENGINEERING

Thursday, August 14, 14
Provenance and the Web

- CrowdLabs: share approaches, results, data, and explorations
- Support visualization, online exploration

www.crowdlabs.org
VisTrails for Teaching Scientific Visualization

• “Using VisTrails and Provenance for Teaching Scientific Visualization” [Silva et al., Eurographics Educator Program, 2010]

• Same features that scientists use for exploratory tasks can also benefit students
  - Exploration: see all pipelines not just a “final” one
  - Comparison: see different pipelines and what changes exist
  - Assessment: see how a solution was developed
Provenance Analysis of Projects

Activity Histograms by Date

The data in the previous section shows that workflow evolution provenance data can help in understanding how different users approach a problem. Figure 5 shows two trees created by different users for the same task. User 1 and User 2. A smaller branching factor indicates that a more direct path was used to obtain a solution. In contrast, a larger branching factor indicates that more trial-and-error was used. Figure 6 shows more structural activity than Tasks 2, 3, and 4. This indicates that, as users have to remember what they did, they would select a previously defined workflow, they would select a tagged node because it is easier to identify. Workflow evolution information can also be helpful to characterize tasks. As noted in Table 4.2.2. Analysis of Tasks, parameter actions involve in workflow; and layout actions involved in workflow. Parameter values in the workflow; and parameter actions in the workflow. Different types of work involved in a task, we classified the actions involved in workflow development into:

- Structural actions
- Parameter actions
- Layout actions

We classified the actions involved in workflow development into:

Different types of work involved in a task, we classified the actions involved in workflow development into:

- Structural actions
- Parameter actions
- Layout actions

This also makes it easier to have adoption in other places. One of the really nice features of the unobtrusive way that VisTrails captures provenance is that there is no extra burden on the user; they can do their work without caring about it. Provenance information, it is possible for one person to see what another person did, and to easily compare their own work to it. This makes it possible for the instructors to share their own work. With the students, who can easily see who the problem was. Instructors can help in understanding how different users approach a problem. Figure 7 shows two trees created by different users for the same task. User 1 and User 2. A smaller branching factor indicates that a more direct path was used to obtain a solution. In contrast, a larger branching factor indicates that more trial-and-error was used.

VisTrails makes it easier to have adoption in other places. One of the really nice features of the unobtrusive way that VisTrails captures provenance is that there is no extra burden on the user; they can do their work without caring about it. Provenance information, it is possible for one person to see what another person did, and to easily compare their own work to it. This makes it possible for the instructors to share their own work. With the students, who can easily see who the problem was. Instructors can help in understanding how different users approach a problem. Figure 7 shows two trees created by different users for the same task. User 1 and User 2. A smaller branching factor indicates that a more direct path was used to obtain a solution. In contrast, a larger branching factor indicates that more trial-and-error was used. Figure 8 shows two trees created by different users for the same task. User 1 and User 2. A smaller branching factor indicates that a more direct path was used to obtain a solution. In contrast, a larger branching factor indicates that more trial-and-error was used.
Provenance Analysis of Projects

Comparing Paths to Solutions for Two Students

Branching Factor of Version Tree

Task 1  Task 2  Task 3  Task 4  Task 5  Task 6

User 1

User 2

Branching Structure for Task 3 of User 1

Branching Structure for Task 3 of User 2
Conclusion

Orchestration  Abstraction  Quality

Reproducibility  Reuse  Exploration
Questions

www.vistrails.org