**eResearch Australasia- Cloud Services for Big Data and Big Data Visualization BoF**

**Microsoft Research Connections**

Rane Johnson, Principal Research Director, Microsoft Research Connections, Redmond, WA, USA, *ranej@microsoft.com*

John Warren, Senior Research Program Manager, Sydney, Australia, *jowarren@microsoft.com*

**DESCRIPTION**

Summary: In this session learn about three cloud based tools to help researchers, ChronoZoom, Academic Search and Layerscape. In this session we would like to present these tools, discuss what can be used today, where hope to go with these tools in the future, understand from you what you need as researchers for these to help you better in your research and what needs to be revised and how you can be involved with the projects today. All these tools are free and created for the academic community and in partnership with academic researchers.

**ChronoZoom** is an open-source community project dedicated to visualizing the history of everything in the cloud. As such, it seeks to bridge the gap between the humanities and sciences and to enable a nearly inexhaustible repository of readily understandable and easily navigable information. By using Big History as the storyline, we hope to achieve a unified, interdisciplinary understanding of the history of the cosmos, Earth, life, and humanity, enabling users to understand the history of everything.

A cloud-based user experience, **Layerscape** employs powerful, everyday tools to analyze and visualize complex Earth and oceanic datasets—enabling scientists to gain environmental insights into Earth. Users can create and share 3-D virtual tours based on their discoveries and collaborate with the Earth-science community in ways that previously seemed impossible. Build your own virtual tours and experience the possibilities.

Learn about **MS Academic Search** a free academic search engine and a test-bed for new research ideas in related fields such as data mining, named entity disambiguation, and visualization. Users can quickly find out information about researchers, papers, conferences, journals, organizations and keywords. Users can also discover relationships among these entities via various visualizations. As of May, 2012, Academic Search has indexed 38 millions publications and 18 million authors.

**Desired Outcome:**
- Share tools and get more usage from researchers from Asia and Australia
- Understand what needs to be modified for these tools to be more helpful
- Get more researchers involved in evolving the ChronoZoom Tool
- Get researchers to test Layerscape and MS Academic Search and help us with next phase of developments

**OUTLINE**

BoFs can take a variety of forms but often include one or more short presentations followed by open discussion. Please provide a brief outline of the BoF using the following format as a guide.

1. **First topic.** ChronoZoom & Layerscape, Rane Johnson: In this session learn about two cloud based services, ChronoZoom and Layerscape. ChronoZoom is an open-source community project dedicated to visualizing the history of everything in the cloud. As such, it seeks to bridge the gap between the humanities and sciences and to enable a nearly inexhaustible repository of readily understandable and easily navigable information. By using Big History as the storyline, we hope to achieve a unified, interdisciplinary understanding of the history of the cosmos, Earth, life, and humanity, enabling users to understand the history of everything.

   A cloud-based user experience, Layerscape employs powerful, everyday tools to analyze and visualize complex Earth and oceanic datasets—enabling scientists to gain environmental insights into Earth. Users can create and share 3-D virtual tours based on their discoveries and collaborate with the Earth-science community in ways that previously seemed impossible. Build your own virtual tours and experience the possibilities.

2. **Second topic.** Academic Search, John Warren: Learn about MS Academic Search a free academic search engine and a test-bed for new research ideas in related fields such as data mining, named entity disambiguation, and visualization. Users can quickly find out information about researchers, papers, conferences, journals, organizations and keywords. Users can also discover relationships among these
entities via various visualizations. As of May, 2012, Academic Search has indexed 38 millions publications and 18 million authors.

3. **Discussion.** Help us better understand how to better connect with the eResearch community in Australia & Asia to all solutions. Help us determine what features and capabilities are needed to be more of a useful resource. Help evolve the researcher needs in a tool like ChronoZoom and Layerscape. Join the next phase of ChronoZoom development and research.

**SHORT ABSTRACT:**

In this session learn about three cloud based data visualization research projects at Microsoft Research Connections that can be utilized by the eResearch Community.

1. **ChronoZoom** is an open-source community project dedicated to visualizing the history of everything in the cloud. As such, it seeks to bridge the gap between the humanities and sciences and to enable a nearly inexhaustible repository of readily understandable and easily navigable information. By using Big History as the storyline, we hope to achieve a unified, interdisciplinary understanding of the history of the cosmos, Earth, life, and humanity, enabling users to understand the history of everything.

2. A cloud-based user experience, **Layerscape** employs powerful, everyday tools to analyze and visualize complex Earth and oceanic datasets—enabling scientists to gain environmental insights into Earth. Users can create and share 3-D virtual tours based on their discoveries and collaborate with the Earth-science community in ways that previously seemed impossible. Build your own virtual tours and experience the possibilities.

3. **MS Academic Search** a free academic search engine and a test-bed for new research ideas in related fields such as data mining, named entity disambiguation, and visualization. Users can quickly find out information about researchers, papers, conferences, journals, organizations and keywords. Users can also discover relationships among these entities via various visualizations. As of May, 2012, Academic Search has indexed 38 millions publications and 18 million authors.

In this session we will give a brief overview of the tools and goals we hope to accomplish. We would love to discuss with the research community questions to understand eResearch Big Data and Cloud Needs, understand how these tools need to evolve and how researchers would be interested in participating. We would really like to have deep discussion on creating a research tool that bridges the gap between science and humanities to build better connections, insights and researcher collaborations.

**EXTENDED ABSTRACT**

In this session learn about three cloud based tools to help eResearchers, ChronoZoom, Academic Search and Layerscape. In this session we would like to present these tools, discuss what can be used today, where hope to go with these tools in the future, understand from you what you need as eResearchers for these to help you better in your research and what needs to be revised and how you can be involved with the projects today. All these tools are free and created for the academic community and in partnership with academic researchers.

**IN THE FOLLOWING SECTIONS WE WILL GIVE A DEEPER OVERVIEW OF THE THREE RESEARCH PROJECTS AND TOOLS:**

**ChronoZoom:**

ChronoZoom was developed to make time relationships between different studies of history clear and vivid. In the process, it provides a framework for exploring related electronic resources. It thus serves as a “master timeline” tying together all kinds of specialized timelines and electronic resources, and aspires to bridge the gap between humanities and the sciences and to bring together and unify all knowledge of the past.

You can browse through history on ChronoZoom to find data in the form of articles, images, video, sound, and other media. ChronoZoom links a wealth of information from five major regimes that unifies all historical knowledge collectively known as Big History.

By drawing upon the latest discoveries from many different disciplines, you can visualize the temporal relationships between events, trends, and themes. Some of the disciplines that contribute information to ChronoZoom include biology, astronomy, geology, climatology, prehistory, archeology, anthropology, economics, cosmology, natural history, and population and environmental studies.
This project has been funded and supported by Microsoft Research Connections in collaboration with University California at Berkeley and Moscow State University.

We envision a world where scientists, researchers, students, and teachers collaborate through ChronoZoom to share information via data, tours, and insight. Imagine a world where the leading academics publish their findings to the world in a manner that can easily be accessed and compared to other data. Imagine a tool that allows teachers to generate tours specific to their classroom needs.

We will focus on community development of features, capabilities, and content. We plan to work with two communities:

• One community will consist of content providers, humanities and science researchers, think tanks, ischools, and organizations that have digital libraries, digital content, and cultural content. The goal is that the content will be viewed in a time-based tool that will help the history of everything come to life by uniting humanities and science stories. It will also enable data comparisons, timeline comparison and a personal canvas area to enable researchers to do deeper analysis, collaboration and cause/effect relationships.

  Content examples: history of particle physics; history of chemical reactions; climate change; history of the Nile, which could include cultural and scientific data; and the history of Polynesian culture.

• The other community will consist of members of computer science departments and ischools who are interested in big data, data visualization, database design, taxonomy, and informatics. This community will help us build out the next set of features for ChronoZoom to support the needs of researchers and professors.

Big Questions: Some big questions we are trying to answer with ChronoZoom:

• How do you organize huge amounts (terabytes and more) of different types of data (such as audio, video, text, PDF files, and images) logically and so they’re easily consumable?
• How do you retain precision while displaying historical information ranging from billions of years ago to today on one scale?
• How do you prioritize content when you have multiple items for the same time period? For example, an Egyptian history timeline, a Chinese cultural history timeline, an Islamic history timeline, a Japanese-American history timeline, and a particle physics timeline might all share a significant event on the same date in the year 1890—how do you display this on the timeline canvas?
• How do you efficiently draw elements on the canvas by using HTML Document Object Model (DOM) graphical elements and not sacrifice CPU usage?
• How do you ensure the same experience on multiple devices, operating systems, and browsers, so that users will have the exact same experience whether they use a Mac, PC, iPad, iPhone, Windows Phone, or Android device?
• How do you make a third-party authoring tool that incorporates an editorial board approval process and provides online journal annotations to populate the timeline—enabling a new type of peer review of entries?
• Do you require an author to build the timeline and create the story line, or do you automate the pull of data from all creative-commons, freely available datasets and libraries on the Internet?
• How do you compare multiple timelines and data sets?
• How do you retain the ambiguity of history? There are many interpretations of history, and much disagreement among experts about historical events; how do you ensure that students of history have access to the diverse historical information so that they can analyze it and develop their own interpretation of what happened?
Potential Future Features

Some of the possible future features could enable the user to:

- Create personal canvas/timeline/tours
- Generate internal user bookmarks
- Generate a chart dynamically and place it where they want on the timeline
  - Display curve and segmented line graphs and plot events coded for magnitude
  - Phylogenetic trees
  - Svg drawing
- Filter exhibits based on subject
- Choose data from data library
- Customize time direction up to down, down to up, left to right, right to left
- Compare data and timelines
- Share timelines or tours with others via social networking
- Display the uncertainty of dates (approximate dates)
- Show a time range in addition to a specific date in time
- Present multiple interpretations
- Display geo-spacial data

Collaborate

Microsoft Research Connections and the ChronoZoom Project promote the establishment of relationships with academics and other partners (like publishers, broadcasters, NGOs, foundations, and organizations that have rich digital media content that crosses the sciences and humanities) to create valuable user experiences with this new technology. As data gets added to the platform, others can share access to enrich their experiences as well.

LAYERSCAPE:

One of the important technical challenges for the environmental scientist is managing the flow and the visualization of research data. Layerscape harnesses your PC’s graphics processor to visualize large amounts of data—in space and in time. Layerscape could be used to render 3-D visualizations from such diverse data sets as historical surface temperature measurements, chlorophyll concentration, seismic activity, greenhouse gas diffusion, sea ice extent, wind patterns, ocean pH, insect biodiversity, aquifer storage, geothermal heat flux, antelope migratory patterns, or the transport of Saharan dust as it fuels plankton blooms across the surface of the Atlantic Ocean with nitrogen and iron. Layerscape can also create abstract visualizations where you do not necessarily need latitude and longitude—just coordinate axes.

In addition to rendering data in 3-D space and in time, Layerscape has what we call freedom of perspective and free narrative. You can place your virtual eye anywhere you like and connect a sequence of perspectives and automated transitions that emphasize what the data is doing and what story you want to communicate. Such storytelling is ideal for educational outreach, enabling you to share your results with the scientific community and the general public.

Layerscape consists of three parts. Part one is the WorldWide Telescope visualization engine, and part two is the website that supports communities of users and the content they (you!) generate. The third part is a tool for getting data into Layerscape. This tool is built on Microsoft Excel, so if your data is already in an Excel spreadsheet, you simply click a few buttons to send it to the visualization engine. The link from Excel is dynamic, meaning that as you change the data in Excel, your Layerscape rendering changes automatically.
Today, a number of scientists—from geologists to seismologists to oceanographers—are using Layerscape to study atmosphere circulation, validate climate models, and even unravel evolutionary patterns of seahorses, demonstrating the wide applicability of Layerscape. Lee Allison, state geologist and director at the Arizona Geological Survey. His agency plays a critical role in public policy decisions, using science to help keep people and property safe, to promote economic resource utilizations, and to protect the environment.

It’s a job Lee clearly enjoys. “Everything about this job is exciting,” he says. “We’re exploring areas that have never been explored before. We’re doing new things with technology that have never been done before. It’s a chance to explore.”

He adds, “It’s that interplay of being able to go out into the field and look at the rocks, discover things that people have never seen before, bring it back into the office and translate it—to tell people what it means to daily life.”

Lee points out that the age of discovery is alive and well, especially underground. “The subsurface is an area that’s really never been fully explored. And that’s what we’re doing here.”

“Now, we’re mapping the geology in 3-D and through 4-D, through geologic time. We’re doing it by mapping the geology on the ground, but then using technology to go well beyond where we can travel as individuals.”

“And Layerscape is this incredible visualization tool that’s coming along that’s going to allow us to take all of these data that we bring together and view it in 3-D and be able to go in the subsurface and be able to fly around and look at it in ways that we’ve never been able to before. This whole concept of visualizing the data is revolutionizing the way not only we do our science, but the way we portray our science to the people who use that data. Not only the public, but to industry and government decision makers.”

From the ocean depths to the high desert, Layerscape is helping scientists visualize complex data, achieve new insights, collaborate with far-flung colleagues, and explain their work through narratives. Build your own virtual tours and discover the possibilities with **Layerscape**.

---

**ACADEMIC SEARCH:**

Microsoft Research\(^{[2]}\) offers a growing suite of software, tools, and services—freely available to the academic community—to help create, enhance, and extend the reach of your repository system. Our goal is to accelerate the research process through innovative tools and technologies, and we are committed to developing and supporting open access, open tools, open technology, and interoperability.

The Microsoft Research Connections team partners with universities, national libraries, publishers, and governmental organizations to develop tools and services to evolve the scholarly information lifecycle. These projects demonstrate our on-going efforts towards producing next-generation tools and platforms in order to increase productivity and empower researchers to increase the discoverability and appropriate re-use of their work.

**Microsoft Academic Search**

Microsoft’s Bing\(^{[3]}\) service is a comprehensive web search engine that indexes materials from all walks of life. It includes scholarly content, but is by no means limited to it. Academic Search, by contrast,

---

\(^{[2]}\) [http://research.microsoft.com](http://research.microsoft.com)

\(^{[3]}\) [http://www.bing.com](http://www.bing.com)
concentrates exclusively on scholarly materials. This means that not only are all query results relevant to the research community, but also that Academic Search contains features uniquely geared toward this audience. Academic Search also emphasizes an entity-based experience, surfacing the people (authors), places (institutions and conference venues), and things (research outputs) that are relevant to the research audience.

**Ranking and Sorting**
In Microsoft Academic Search, objects in the search results are sorted based on two factors: their relevance to the query and their static rank value. The former denotes the extent to which the returned result meets the information need of a given query. The latter encompasses the authority of the result, including such important details as how often and where a paper is cited. For each subject domain, Academic Search provides a ranked list of publications, authors, journals, institutions, and conferences. End-users can re-sort the lists in several ways. Top-author lists, for example, can be sorted by h-index[^4] or citation count, and the aggregations can be scoped by year ranges. Top-organization lists can be sorted by year range as well as by geographical region.

**Citation Data and Citation Context**
A critical component of academic literature searching is reference linking. To be able to follow the papers cited by an article, and the papers that, in turn, cite it is critical to information discovery. Microsoft Academic Search provides an efficient mechanism to follow this chain. Search results, whenever possible, display lists and links to both referenced and citing papers: In addition, Microsoft Academic Search highlights the context of a citation (when possible) by displaying the section of the citing paper in which it is referenced.

**Author and Institutional Profiles**
Author and institutional profiles are dynamically created within Microsoft Academic Search. These pages contain publication histories, citation information, collaboration details, publication lists, research interests, institutional affiliations, and more. The information associated with author and institutional profiles is derived from the tens of millions of scholarly articles currently indexed within the service.

Most of these articles have reference lists that Microsoft ingests and processes. The indexed articles and ingested references help create a snapshot of individual authors' publication history, productivity, and impact. As more content is indexed within Microsoft Academic Search, the accuracy and completeness of our author profile data will continue to improve.

**Visualization Features**
Microsoft Academic Search also provides a number of elements to visualize research information:

**Paper Citation Graph** – This feature shows the citation relationships among publications, providing an innovative way of navigating from a particular resource to the papers that cite it. Double-clicking a citing article will re-draw the graph, showing the papers that cite that one. And clicking on the edge between any two papers will expose the citation context, provided such information is available in the index.

**Genealogy Graph** – Scientific discovery is a continuum, with the graduate student becoming the professor and, eventually, the mentor. The Genealogy Graph feature captures this continuum by displaying the historical relationships among mentors and their advisees. This tool provides a unique look into the lineage of individual scientists and their research outputs.

**Academic Map** – This feature makes it easy for visually to map scholarly output from organizations geographically. Overlaid upon Bing Maps, the Academic map provides a worldwide view of the

institutions within the index, allowing users to filter by domain and to zoom into geographic locations to locate and view researchers and publications.

**Domain Trend** – The Domain Trend page analyses research trends of academic domains and subdomains, then visualizes this data using a stacked area chart. It provides users with a clear view of how publication interests have changed over time, and displays a list of top authors during a selected time period.

**Relevance for Repositories**

**Programmatic Access to Data** – Microsoft Academic Search has a number of ways that information can be consumed and surfaced within your repository: Embeddable (JavaScript) widgets create a simple way for repositories to dynamically fetch and embed Author profile data within a local repository site. RSS feeds are available for Authors, Journals, Publications, Conferences, Keywords and Organizations.

Microsoft Academic Search also provides an application programming interface (API) that allows repositories and other applications to build compelling tools and experiences on top of our rich data. Whether you want to build your own ranking of institutions in agronomy or build a visual explorer for browsing academic papers in virology, we believe the APIs make it easy for you to start that project. The Academic Search API currently supports both REST/JSON and SOAP/XML options for consuming data from the Academic Search index.

Among the new and innovative uses of Microsoft Academic Search data are the Eigenfactor\(^5\) project, which has been exploring the flow of citations between the subdomains of Computer Science literature and producing new interactive maps, and the ScienceCard\(^6\) project which is aggregating various forms of article-level metrics. We will continue to improve the APIs and we look forward to closely working with the repository community to ensure that the APIs deliver the reliable, high-quality results so that you can focus on your apps and services.

**Crowdsourcing** – Microsoft Academic Search is open for users and institutions to edit and enhance. If you find inaccurate or out-of-date information, corrections or updates can be submitted online. Microsoft will verify requested changes to Microsoft Academic Search. We appreciate contributions to help improve Microsoft Academic Search’s accuracy and completeness.

**Desired Outcome:**
- Share tools and get more usage from researchers from Asia and Australia
- Understand what needs to be modified for these tools to be more helpful to researchers
- Get more researchers involved in evolving the ChronoZoom Tool
- Get researchers to test Layerscape and MS Academic Search and help us with next phase of developments
- Understand what we need to do better for these tools to be utilized in your research work
- Understand what other cloud based big data interests there are and how we can help

**About the Speakers**

Rane Johnson-Stempson is the Education and Scholarly Communication Principal Research Director, where she engages with academics worldwide and identifies high-impact areas for research investigations. She is currently working on projects that use technology to transform how we learn about history and how we eradicate human trafficking. Rane is also the lead for growing, attracting and retaining women in research,

\[^5\] [http://www.eigenfactor.org/](http://www.eigenfactor.org/)
\[^6\] [http://sciencecard.org/](http://sciencecard.org/)
science and engineering. She is actively working with NCWIT, Anita Borg, CRA-W and researchers on how to grow the pipeline of women in research, science and engineering. Previously, as the WW Director of Education Strategy she was responsible for multi-stakeholder partnerships, Cisco-Intel-Microsoft Alliance, World Economic Forum- Global Education Initiative, Fast Track Initiative-Private Constituency and helping our local teams around the world use the power of technology as an accelerator to transform education to meet local government priorities to drive economic and workforce development in the 21st century. Rane is very passionate about education and technology with sixteen years of experience in industry, non-profit and government. In addition, she has sat on the boards of top Education non-profits and on legislative committees and task forces. Rane is a graduate of Bucknell University with a BS Mechanical Engineering, BA Economics/Finance, and George Fox University with an Executive MBA in Transformational Leadership. Her research interests lie in technology implications in reaching education for all, technology and the interplay of public-private partnerships in solving our greatest social problems and technology innovations required to engage student learning and grow women’s participation in STEM.

John Warren is the Manager, Australia and New Zealand Research Programs for Microsoft Research Connections. Based in Sydney, John works with academics and with State and Federal Governments to identify innovative and impactful research projects that complement the research themes developed by Microsoft Research Connections. Microsoft researchers from our labs around the world typically collaborate with ANZ researchers on projects developed around themes determined by Microsoft Research Connections. These themes currently include Health and Wellbeing, Natural User Interface, Earth Energy and the Environment, Computer Science and Education & Scholarly Communications. During John’s more than 30 year’s experience in the IT Industry in Australia, New Zealand and Asia working with blue chip organizations, he has developed an interest in working with leading academics to identify and address significant research challenges. Specifically the past 15 years has been devoted to helping to develop the local research community via internships, PhD scholarships and highly impactful research projects by some of our leading academics. A graduate of the University of New South Wales, with a degree in Psychology, John’s personal research interests lie in exploring the ways in which technology is becoming more natural and intuitive.