THE NEW SCHOOL
68 5th Avenue
Room 200
New York, NY 10011
T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

Design and Visualization Best Practices for Big Data: Enhancing Data Discovery through Improved Usability

Expediting Cooperation in Government-funded Open Source Programs: Incremental Agent-based Mapping, a Pattern Language for Collaborative Cognition

Award No. FA8750-12-2-0325

Bernd Riedel, Chris Goranson
The Parsons Institute for Information Mapping, of The New School

October 14th, 2013

This work was supported by the Defense Advanced Research Projects Agency (DARPA) Award No. FA8750-12-2-0325. Any design guidance, opinions, findings, conclusions or recommendations expressed here are those of the authors and do not necessary reflect the views of DARPA.

THE NEW SCHOOL

68 5th Avenue
Room 200
New York, NY 10011

T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

BACKGROUND

In 2012, the Defense Advanced Research Projects Agency, or DARPA, began a multi-year initiative to make advances in analytics, data visualization and usability solutions for big data. The program made the unique decision to follow a collaboration model more similar in many ways to that of a start-up; enlisting the support of 24 independent performers working across three primary topic areas to organically develop methods and tools in support of the program's objectives. A more-typical format for federally funded government research at such a scale involves one primary government contractor facilitating the work of many other subcontractors across a previously defined schedule, with clear deliverables, performance metrics and dates. Instead, DARPA provided the performers with project-specific deadlines, goals and further guidance, but encouraged teams to develop their own frameworks for success and encouraged collaboration. Because the program had also adopted a stringent open source policy, performers that were normally competitors in the open market had little reason not to work together in order to achieve meaningful results for the program.

The 24 independent performers represented a very diverse cross-section of companies, institutions and academics involved in fields related to big data. Generally divided into three key domain areas (algorithm development, systems/storage, and visualization), and into two primary functional areas, TA1 and TA2. Performers that fell into TA1 tended to provide algorithm development and GPU/database development support – essentially the infrastructure necessary to rapidly analyze, process and store vast amounts of data. TA2 performers on the other hand were a smaller group consisting primarily of companies and institutions that specialized in visualizing data or in the frameworks necessary to present information to a user. Many performers and their assembled teams had very specific project goals that largely dictated their expertise and skills, but very few, if any would have had the ability to cover all three sufficiently to build the next great big data application for broad dissemination by DARPA.

Indeed, as a result this required teams to work together and collaborate; of which the most successful performers at the end of year one appeared to be those teams who had actively sought out collaborative opportunities while offering support to performers that might not have had as strong a background in a particular area. Performers that were in attendance for a majority of the summer workshop also seemed to develop more long-term relationships with other performers and benefit from day-to-day working groups on various challenge problems posed by DARPA management.

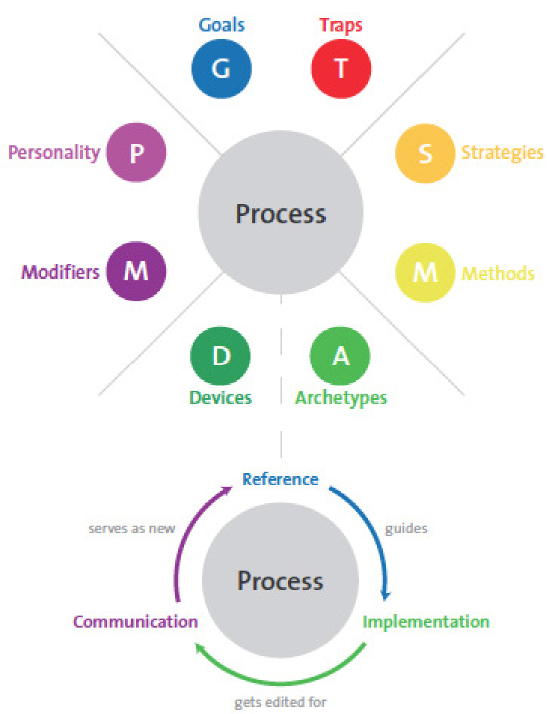
Early on in the project (late 2012), TA2 performers had decided to meet in person at facilities at DARPA to begin planning for the summer workshop. Given the fluid nature of the project and the management structure, one thing that became evident was that there was an inherent need to understand the landscape of each performer's own challenges as they related to overall tasks and responsibilities for their contributions to the program. At the same time, in order to be truly successful going through the summer workshop, performers needed to understand where they ultimately plugged in to the larger program, and begin identifying needs that were universal amongst the teams and identify ways to begin addressing them. This was particularly important when one considers the short time frame of the summer workshop. Had the teams not embarked on this kind of discovery prior to summer camp, many more hours would have been wasted just trying to better understand and isolate opportunities for collaboration with the other performers.

68 5th Avenue
Room 200
New York, NY 10011

T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

In an attempt to understand the unique phenomenon presented by the programmatic approach, PIIM embarked on a project to apply the research of a Visiting Fellow, Bernd Riedel to document and better understand the interrelationships of the teams. Riedel's conceptual mapping is a form of externalizing knowledge through mental visualizations. Called, the Incremental Agent-based Mapping Framework (IAM), this research explores a possible new role that information visualization can take in collaborative processes and complex systems.

THE INCREMENTAL AGENT-BASED MAPPING FRAMEWORK (IAM)



The Incremental Agent-based Mapping framework (IAM) proposes guidelines to assemble and compile individualized mapping experiences in the form of activities, workshops or exercises. For the DARPA program, this meant that through a series of short working sessions teams would be responsible for self-identifying their own specialties and those of their partners. In a sense, this is similar to a competitive analysis that any one of the performers might have done on their own; the difference here is that given the open nature of the project, performers were self-identifying their own strengths and weaknesses (as well as those they perceived of the other teams) in a public way. Following an IAM facilitator, an incremental set of actions was used to navigate the participants through the complexity of a design process to better identify these interrelationships. This aspect of diverse perspectives, the *'personal voice'*, often gets lost during codification (simplification and standardization) in traditional mapping methods, but became the structure on which the various performers began to realize and subsequently document their contributions to the project.

Instead of simply documenting the performer's interactions through a map (e.g. infographic), IAM focuses primarily on capturing the process of discovery the various performers follow. Mapping is therefore seen as a process that allows for the ongoing negotiation of multiple perspectives and value systems. IAM embraces the expressive cognitive style of each mapmaker, while providing appropriate mapping components on the spot for the facilitator. It explores the possibilities of an agile framework for incremental collaborative conceptual visualizations, or maps, while relying not on improved technology or digital tools to aid this process, but rather a facilitated process of information transfer and simple intuitive mapping exercises. This approach turned out to work extremely well as many of the TA2 performers had diverse backgrounds from each other, making one universal approach potentially limiting in effectively documenting each contribution. Viewing the outcomes not as maps or diagrams, but rather as visualizations allows a for a new way to formulate and harvest information, and a different way to navigate complex projects like that presented by the DARPA XDATA program.

68 5th Avenue
Room 200
New York, NY 10011
T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

INCREMENTAL AGENT-BASED MAPPING

The IAM framework really represented in many ways an agile mapping technique. The effective use of the framework requires the participation and input of multiple agents, who are all participants in a process that is subject to analysis. The exercises in the workshop are incremental in nature, meaning that each subsequent round is designed differently and according to revealed agent's inside knowledge, adding or subtracting a certain variable based on the findings of the previous workshop, and observing how this alters the outcome. Because all evaluations are done openly in a workshop format, performers have the opportunity to see how they are viewed as overall contributors to the end-goal, and are encouraged to openly discuss discrepancies and opportunities.

To facilitate a mapping activity effectively, a facilitator must understand the differences between relevant mapping tools (methods and strategies), while being able to contextualize and identify conditions of a mapping activity. IAM should identify both opportunities as well as potential roadblocks to overall program success through the eyes of each performer. A precise knowledge on the *'DNA'* elements of mapping (devices and archetypes) is important for both reading and enhancing maps that are produced by participants, as well as an understanding of what drives the activity. This includes understanding the program well enough to capture both external factors that can be actively addressed by the facilitator, as well as the internal *'mind-set'* of the participant.

As a guideline to facilitate these activities, the IAM framework offers a simple iterative process to:

- Create the conditions for knowledge transactions (the *'reference'*)
- Make it happen (the *'implementation'*)
- Document and process the outcomes for the next cycle (the *'communication'*)

IAM was developed following extensive literature reviews, as well as experiences in the applied multi-stakeholder design projects of the experimental MFA Transdisciplinary Design program at Parsons The New School for Design. The IAM approach was further field-tested in the XDATA program TA2 group research phase between February and May 2013.

THE APPLIED FRAMEWORK OF IAM

Reference Phase: Conducting a Data-Dive with Program Performers

First, the goal of creating a *'performer landscape'* is to begin to visualize the many divergent aspects of the program– in this case XDATA. This establishes the reference phase by creating basic performer profiles based on research and surveys. This results in an incomplete view where each performer is concerned, since each performer is only sufficiently able to view a limited aspect of the whole program. Early on, for example, TA2 performers struggled not only to understand what other TA2 performers were offering, but what TA1 performers were. Since each performer was offering, in many cases, a set of highly specialized assets, understanding how all of these could eventually fit together was akin to solving a giant jigsaw puzzle without the advantage of having the final picture to look at. Of course, most performers relished the opportunity to define what the final picture would actually look like, so each was sufficiently motivated to cooperate to help build on the overall vision of the program.

68 5th Avenue
Room 200
New York, NY 10011

T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

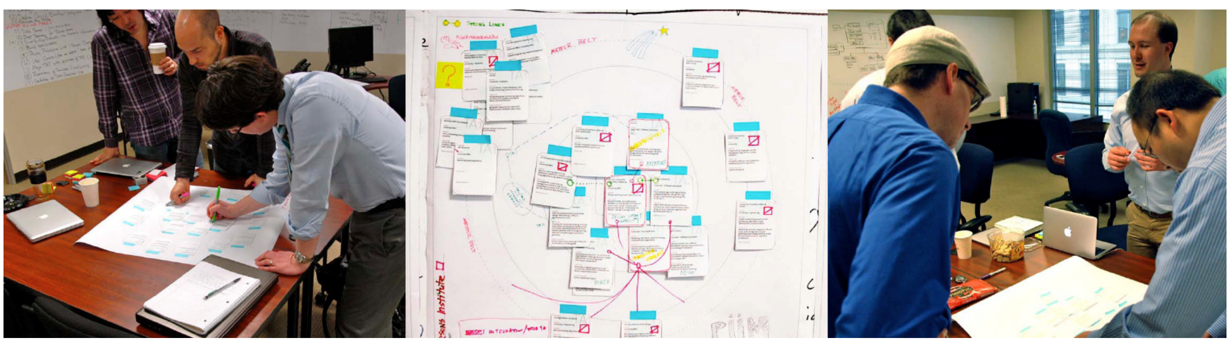
These findings were then presented back to the performers through the XDATA wiki in the form of a comprehensive list consisting of the participating entity's name, self-described expertise, project goals and existing partnerships within the project. Actors completed and added content to the list, while extending it with a first emergent topic: a software component list.

Implementation Phase: Conceptual mapping at the TA2 Hackathon

Roughly 30 persons representing seven companies and universities of the TA2 group participated during a hackathon held April 17-18th, 2013. Two individualized mapping activities were conducted to gain a collaborative overview on workflow issues.

Mapping Activity 1: Performer Relation Maps

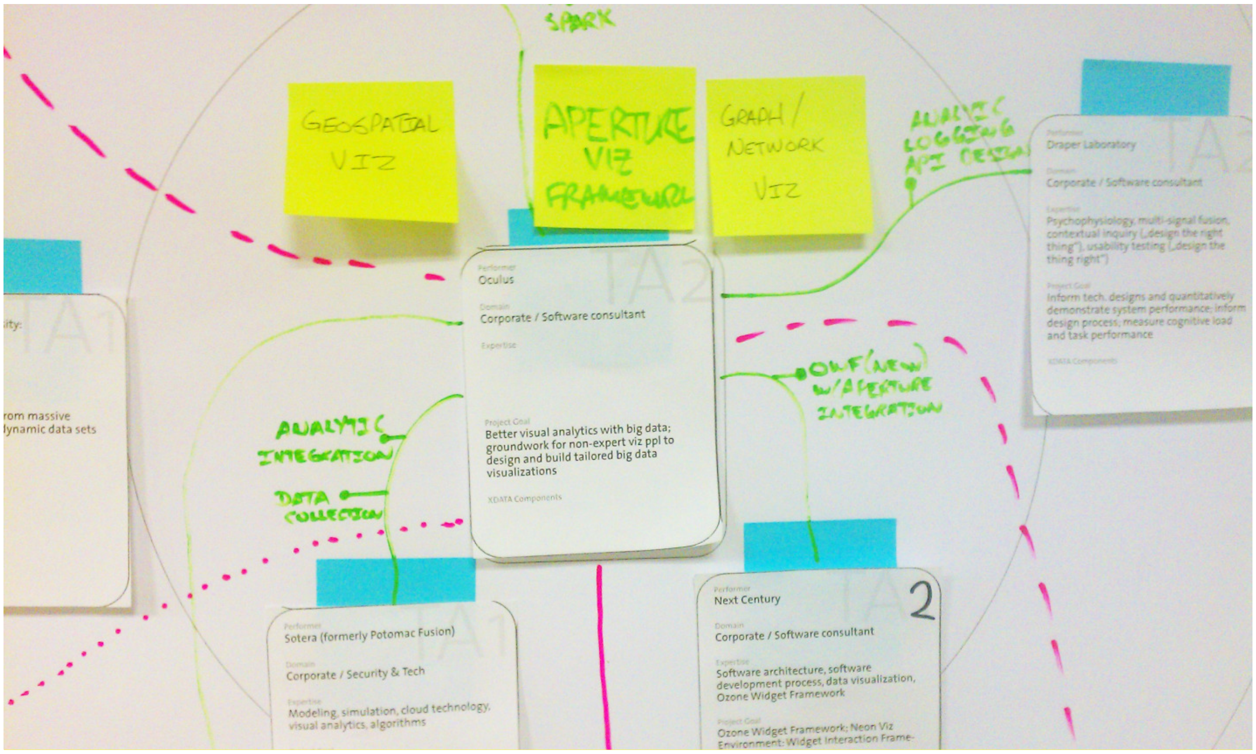
Performer Relation Maps were designed to illuminate how a person or entity in a complex system is seen by other stakeholders in the same system. The facilitator creates *'performer profiles'* in the form of flash cards based on information retrieved in the reference phase. Every participant receives a full set and then arranges the cards they deem most relevant in a provided circular template with their own card in the center, and the other cards in relation to how *'close'* they see the other participants. Connections are drawn and labeled and additional information (e.g. shared resources, skills or interests) is added on post-its. The individual maps are then displayed to all participants and each participant can see his or her relation to specific other actors through their eyes. This opens opportunities for re-evaluating how actors see themselves and how others see them.



These excercises helped identify some themes that proved to be predictive throughout the Summer Workshop. One example can be found in the Oculus map, pictured below. In it, Draper's logging API and Next Century's Open Widget Framework / Neon assets are both identified. Both of these assets were instrumental in the final summer workshop presentations of all three groups; Draper's logging API was used to collect instrumented feedback on the Oculus interface, and Neon deployed a widget using one of the Oculus visualizations for a formal demonstration. Early planning and brainstorming sessions like those held by the TA2 performers through this method helped performers realize and lay the groundwork for opportunities to work together.

68 5th Avenue
Room 200
New York, NY 10011

T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu



Mapping Activity 2: Collaborative Mapping Exercise

A Collaborative Mapping Exercise was held throughout the whole hackathon. During designated short intervals, mostly working breaks and small talk occasions, changing groups worked on one large-scale process map (60 x 120") template, to map out relevant components for their cooperation in the XDATA program.

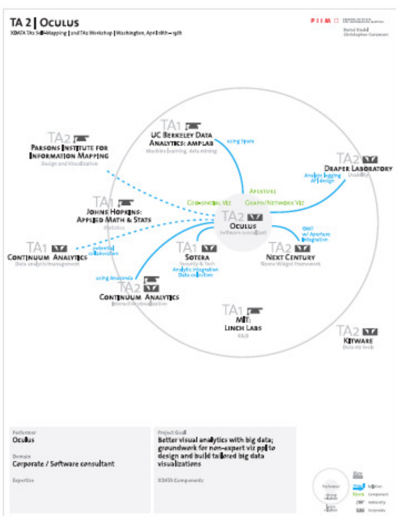
This way pressing topics were identified (e.g., indispensable software components), responsibilities collaboratively distributed (e.g., who takes ownership of each component), relations between performers extended (e.g., development processes were shared/divided), and logistical problems identified and addressed (e.g., existing work flow barriers). During this process, the map served as a *'visual agreement'* and as a reference for ongoing discussions. The image above identifies tasks along a timeline that represent both opportunities

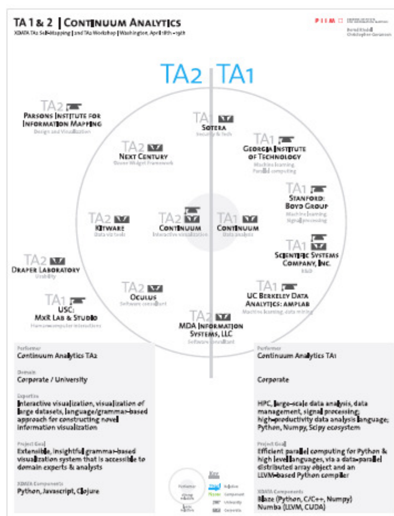
68 5th Avenue
Room 200
New York, NY 10011
T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

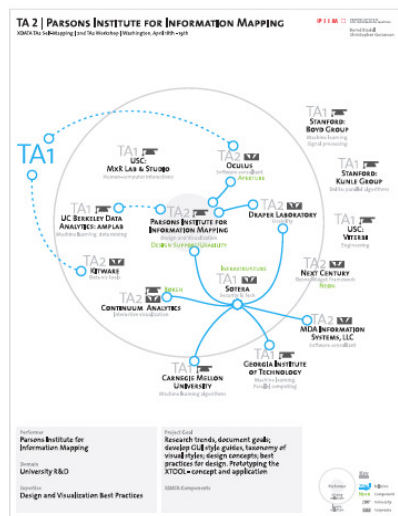
and challenges ahead of the summer workshop. By identifying dependencies early on, the performers were able to position themselves to address core challenges presented by DARPA's program managers.

Communication Phase: Processing and display of outcomes

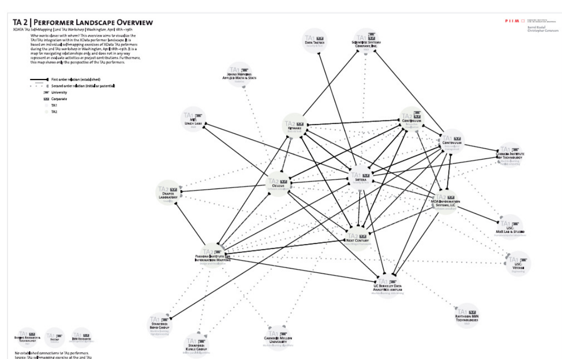
Following the IAM process of reference - implementation - communication, visualizations were created to feedback the outcomes to all performers through the XDATA wiki. Each performer's map identified connections amongst other performers and were used to help solidify approaches to work during the summer workshop. The weakness of TA1 representation is also telling; because we did not have access to those teams through the various planning activities, performers struggled in many cases to identify partnership opportunities. During the first few weeks of the summer workshop, performers had to spend additional time understanding the complexities of new performer's projects while attempting to identify partnering possibilities. By the mid-point of the summer session many these relationships were better established, but it's likely that application of the IAM methodology to both TA1 and TA2 performers would have been helpful in establishing a better foundation for collaboration.







A performer overview was created based on an average derived from the individual performer relation maps. It aims to visualize the TA1/TA2 integration within the XDATA performer landscape. It is a map for navigating relationships from the perspective of the TA2 performers. It does not however attempt to evaluate activities or project contributions. Please see the Appendix A and Appendix B for the TA2 Performer Landscape Overview and the individual team maps.



THE NEW SCHOOL
68 5th Avenue
Room 200
New York, NY 10011
T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

GOALS AND OBJECTIVE: COLLECTIVE COGNITION

IAM is a way to use mapping to explore complex systems from a perspective other than traditional concept mapping or mind mapping exercises. It does this by including multiple perspectives in a process that strongly promotes participation of all parties involved without preconditions in skill or mapping experience, thus shifting the traditional power structure of professional map production. The result is the achievement of an inclusive process of knowledge production. A framework to compile individualized mapping exercises in a time-efficient manner can be achieved by choosing components of the IAM framework in alignment with the intent for using mapping, following an incremental and iterative process based on information visualization (reference-implementation-communication) that aids the designer in organizing information flow and workflow. The purpose is to extract information and knowledge that is otherwise difficult to access or communicate, and even information that a participant might not be aware of knowing.

CONCLUSION

We believe that IAM may be an effective way to encourage and enhance participation in government-funded research that require a significant amount of collaboration amongst performers who under normal circumstances might see each other as competitors. Given the rather unique format of the DARPA XDATA program we think that this may serve as a model for documenting teaming opportunities, program dependencies and help teams better prepare for time-intensive hackathons. We believe that IAM may help to build a significant level of trust amongst performers, but probably is most effective when all performers are equally motivated to produce open source deliverables.

This framework is merely a first attempt to explore these knowledge conditions in detail. It serves both as a wireframe for a library of tools and activities and can serve as a basis for an instruction manual for creating collaborative mapping experiences for participants of a creative research project. Because an open source approach inherently relies on effective collaborations, IAM may be a model for DARPA program management to consider enhancing such approaches for future government-funded efforts following a similar model to that of the XDATA program.

THE NEW SCHOOL
68 5th Avenue
Room 200
New York, NY 10011
T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

APPENDIX A

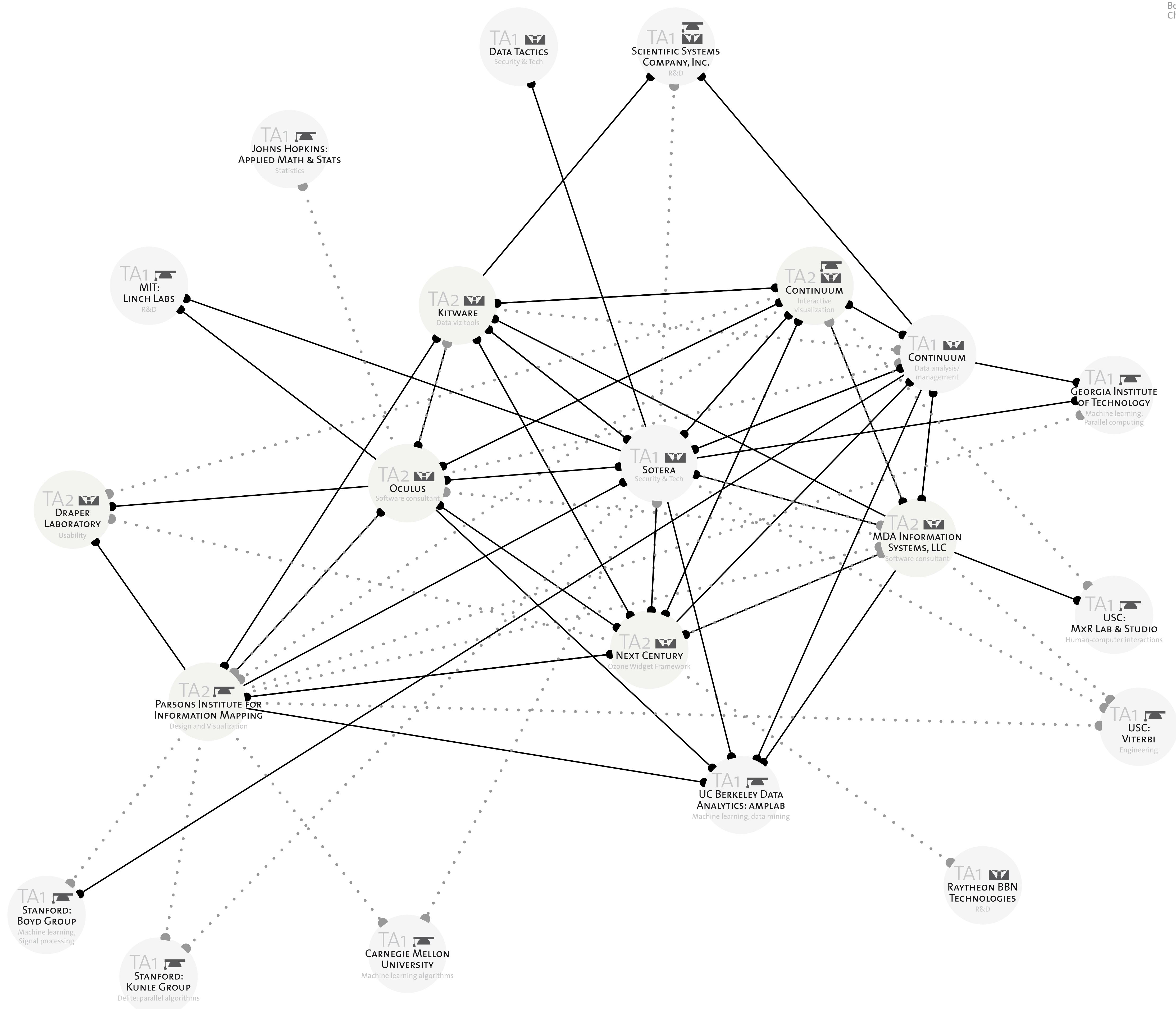
TA 2 | Performer Landscape Overview

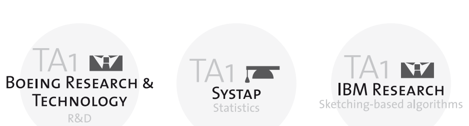
XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM Parsons Institute for Information Mapping
Bernd Riedel
Christopher Goranson

Who works closer with whom? This overview aims to visualize the TA1/TA2 integration within the XData performer landscape. It is based on individual self-mapping exercises of XData TA2 peformers during the 2nd TA2 workshop in Washington, April 18th -19th. It is a map for navigating relationships only, and does not in any way represent or evaluate activities or project contributions. Furthermore, this map shows only the perspective of the TA2 performers.

- First order relation (established)
- Second order relation (initial or potential)
- University
- Corporate
- TA1
- TA2





No established connections to TA2 performers.

Source: TA2 self-mapping exercise at the 2nd TA2 workshop

THE NEW SCHOOL

68 5th Avenue
Room 200
New York, NY 10011

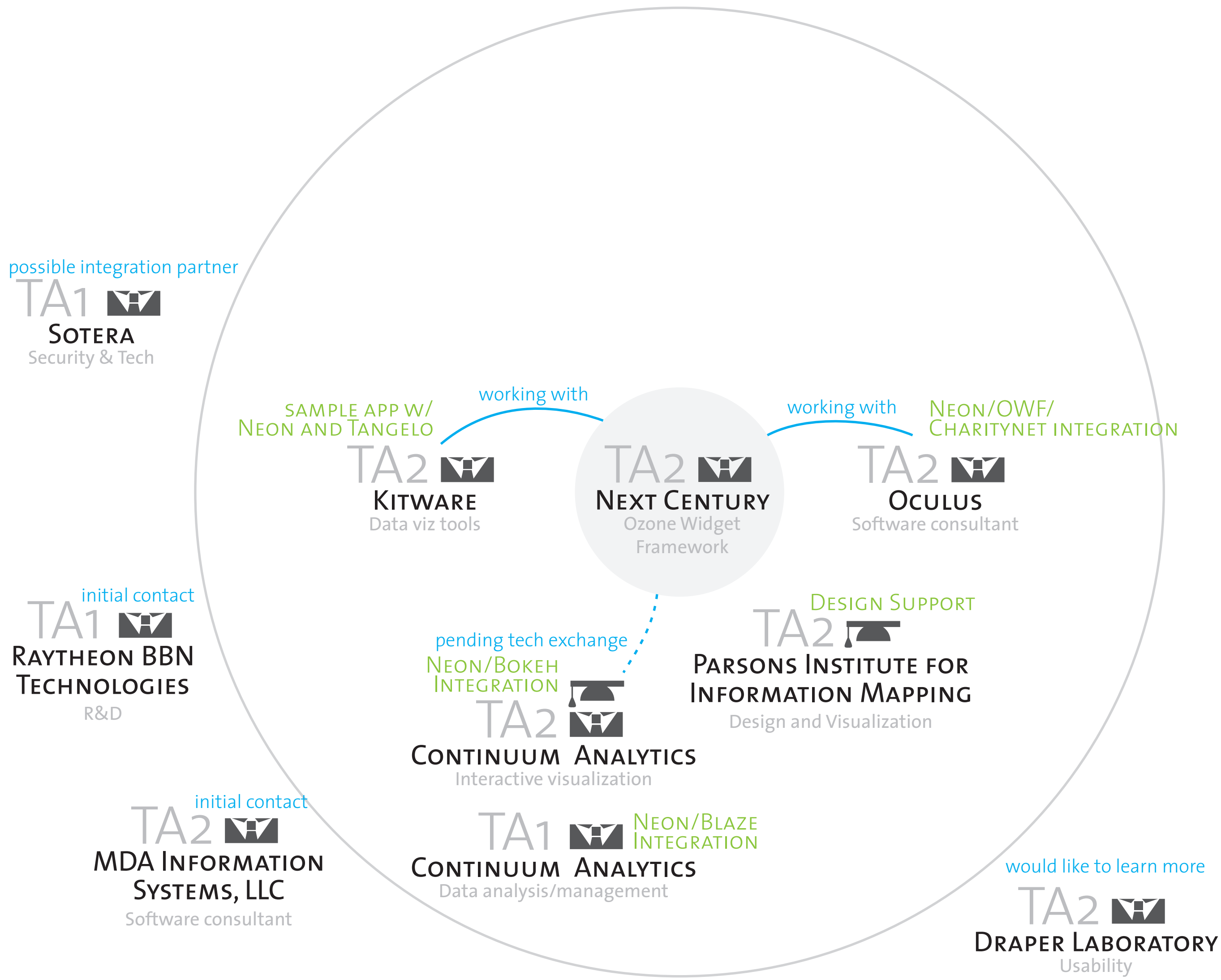
T: 212 229 6825
F: 212 414 4031
http://piim.newschool.edu

APPENDIX B

TA 2 | Next Century

XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM Parsons Institute for Information Mapping
Bernd Riedel
Christopher Goranson



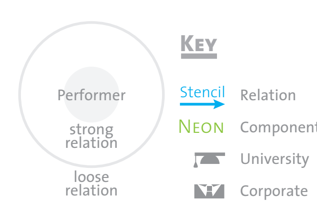
Performer
Next Century

Domain
Corporate / Software consultant

Expertise
Software architecture, software development process, data visualization, Ozone Widget Framework

Project Goal
Ozone Widget Framework; Neon Viz Environment: Widget Interaction Frame-work, Widget Dev

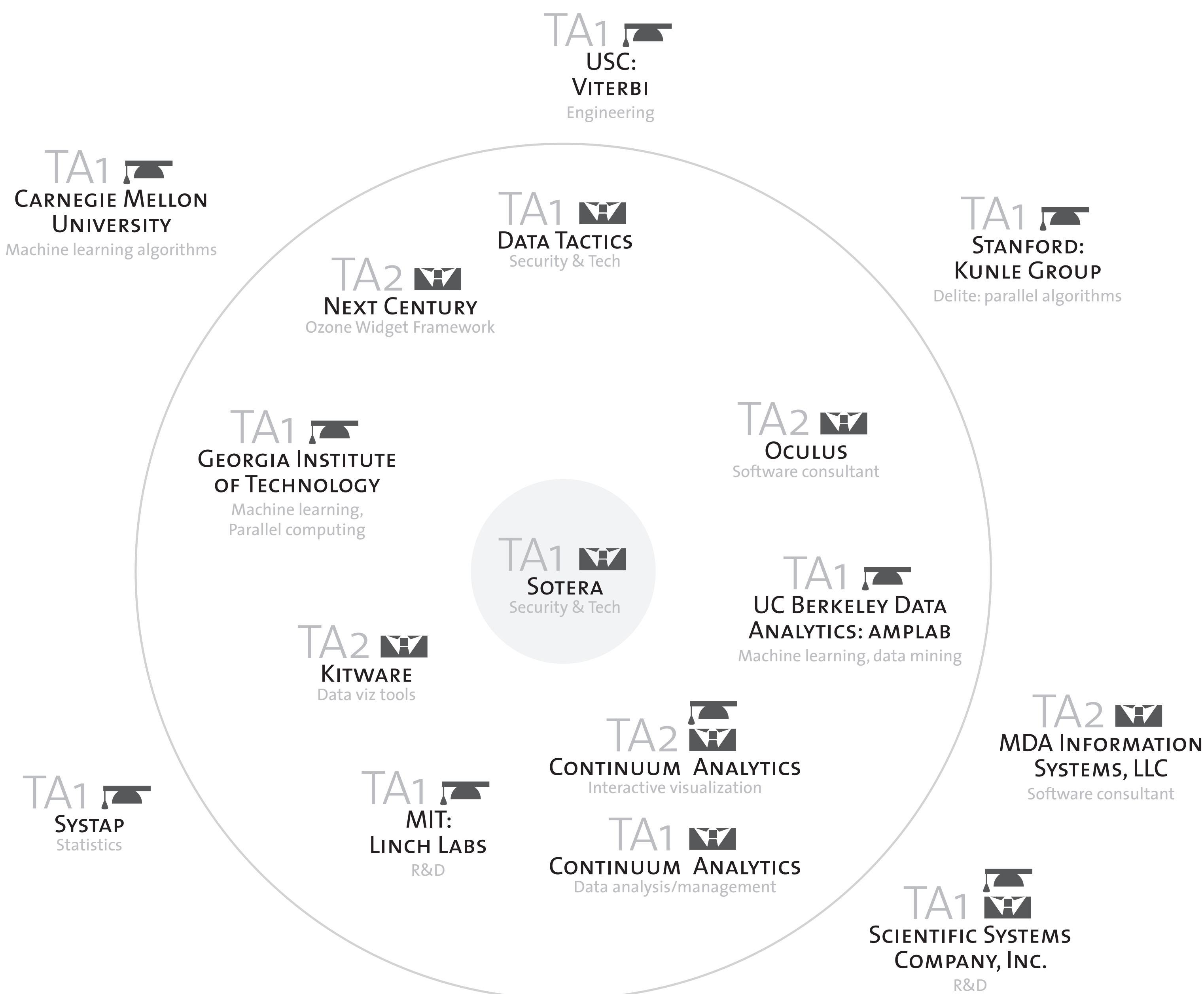
XDATA Components
MongoDB for data storage
Groovy for neon server
Javascript for data visualization



TA 1 | Sotera

XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM Parsons Institute for Information Mapping
Bernd Riedel
Christopher Goranson

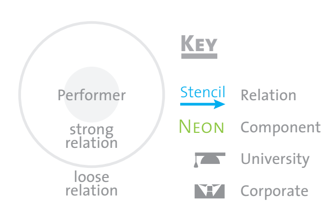


Performer
Sotera (formerly Potomac Fusion)

Domain
Corporate / Security & Tech

Expertise
Modeling, simulation, cloud technology, visual analytics, algorithms

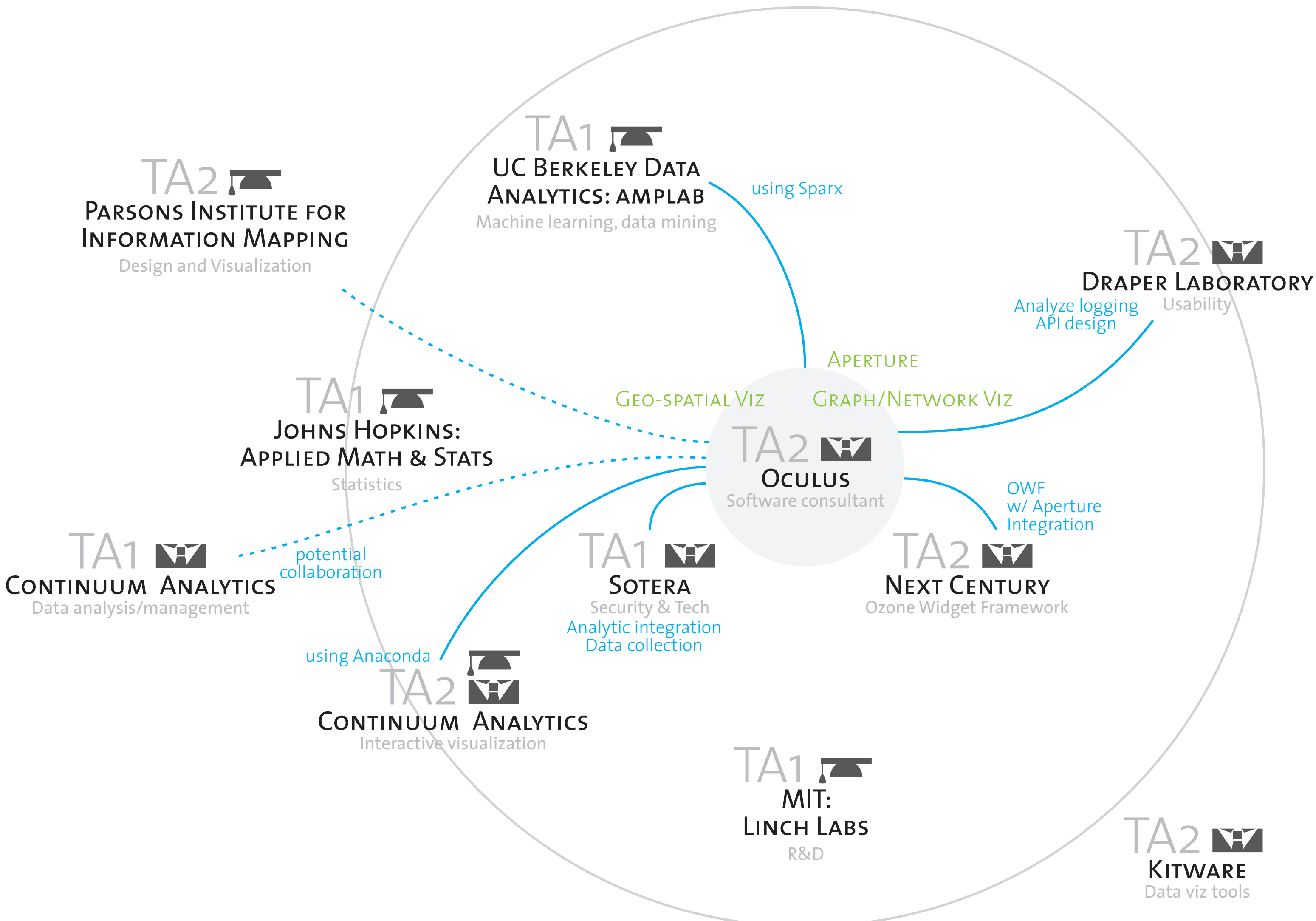
Project Goal
Dimensional and graph analytics; abstraction layer around assymmetric comp env.; OS libraries; quality measurement



TA 2 | Oculus

XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM Parsons Institute for Information Mapping
Bernd Riedel
Christopher Goranson



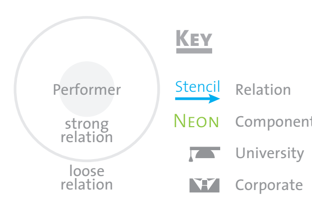
Performer
Oculus

Domain
Corporate / Software consultant

Expertise

Project Goal
Better visual analytics with big data; groundwork for non-expert viz ppl to design and build tailored big data visualizations

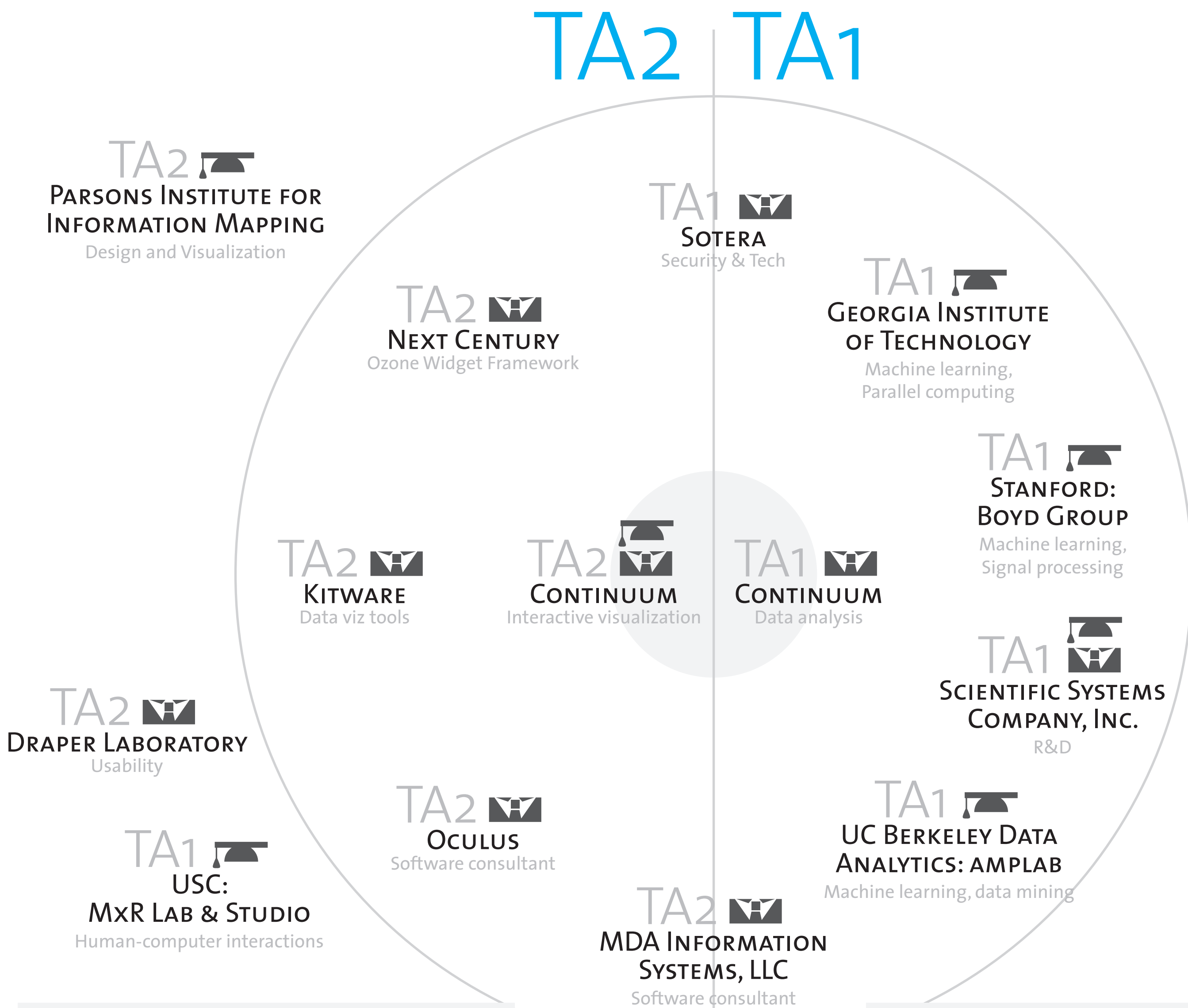
XDATA Components



TA 1 & 2 | Continuum Analytics

XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM PARSONS INSTITUTE FOR INFORMATION MAPPING
Bernd Riedel
Christopher Goranson



Performer
Continuum Analytics TA2

Domain
Corporate / University

Expertise
Interactive visualization, visualization of large datasets, language/grammar-based approach for constructing novel information visualization

Project Goal
Extensible, insightful grammar-based visualization system that is accessible to domain experts & analysts

XDATA Components
Python, Javascript, Clojure

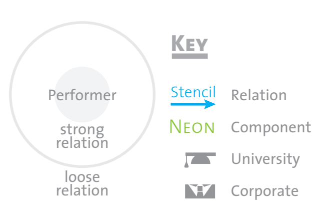
Performer
Continuum Analytics TA1

Corporate

HPC, large-scale data analysis, data management, signal processing; high-productivity data analysis language; Python, Numpy, Scipy ecosystem

Project Goal
Efficient parallel computing for Python & high level languages, via a data-parallel distributed array object and an LLVM-based Python compiler

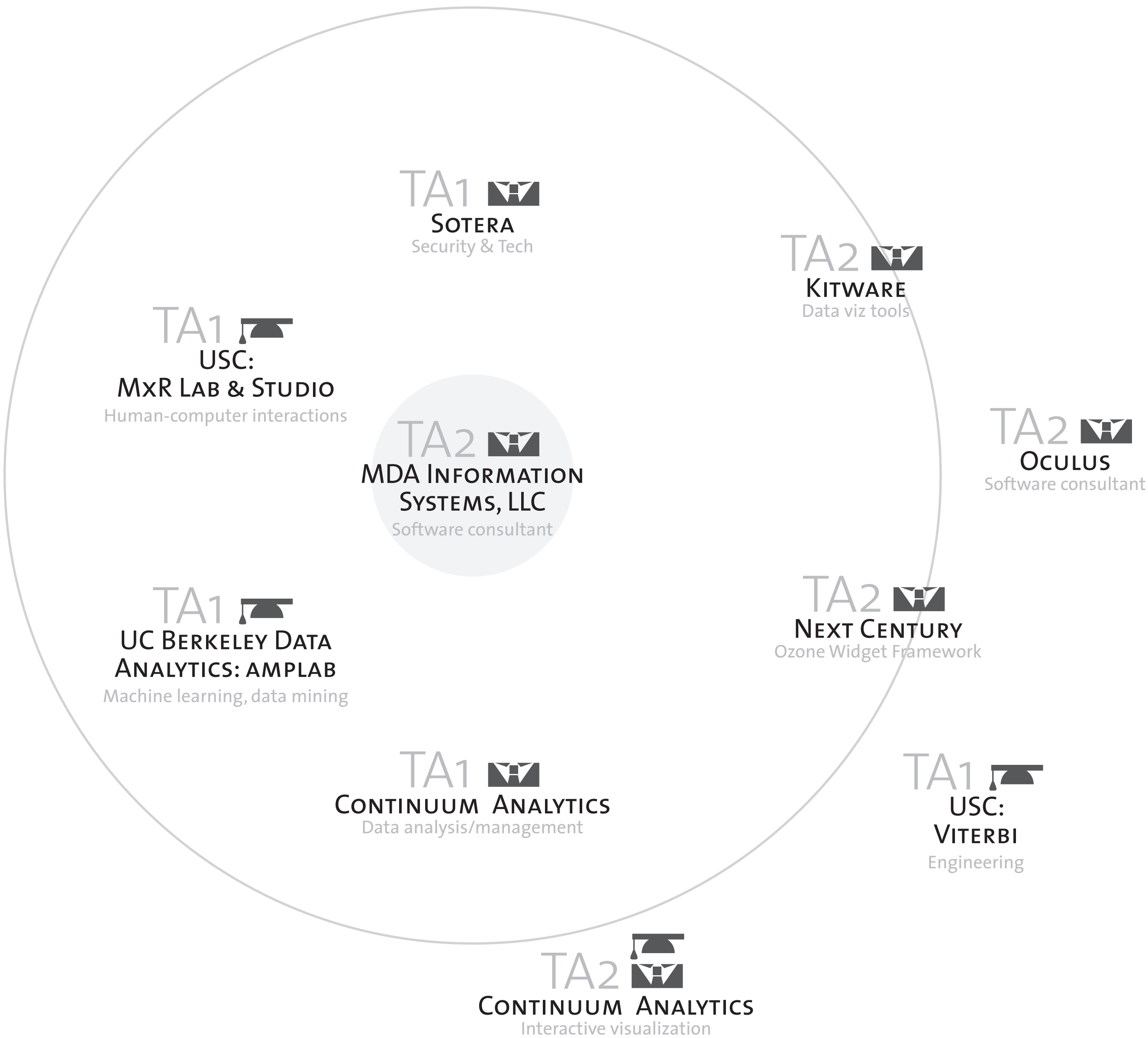
XDATA Components
Blaze (Python, C/C++, Numpy)
Numba (LLVM, CUDA)



TA 2 | MDA Information Systems, LLC

XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM Parsons Institute for Information Mapping
Bernd Riedel
Christopher Goranson



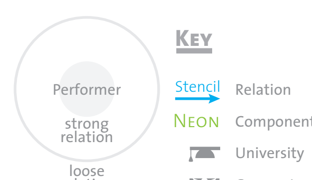
Performer
MDA Information Systems, LLC

Domain
Corporate / Software consultant

Expertise

Project Goal
Unstructured text analysis; sentiment assessment and tracking

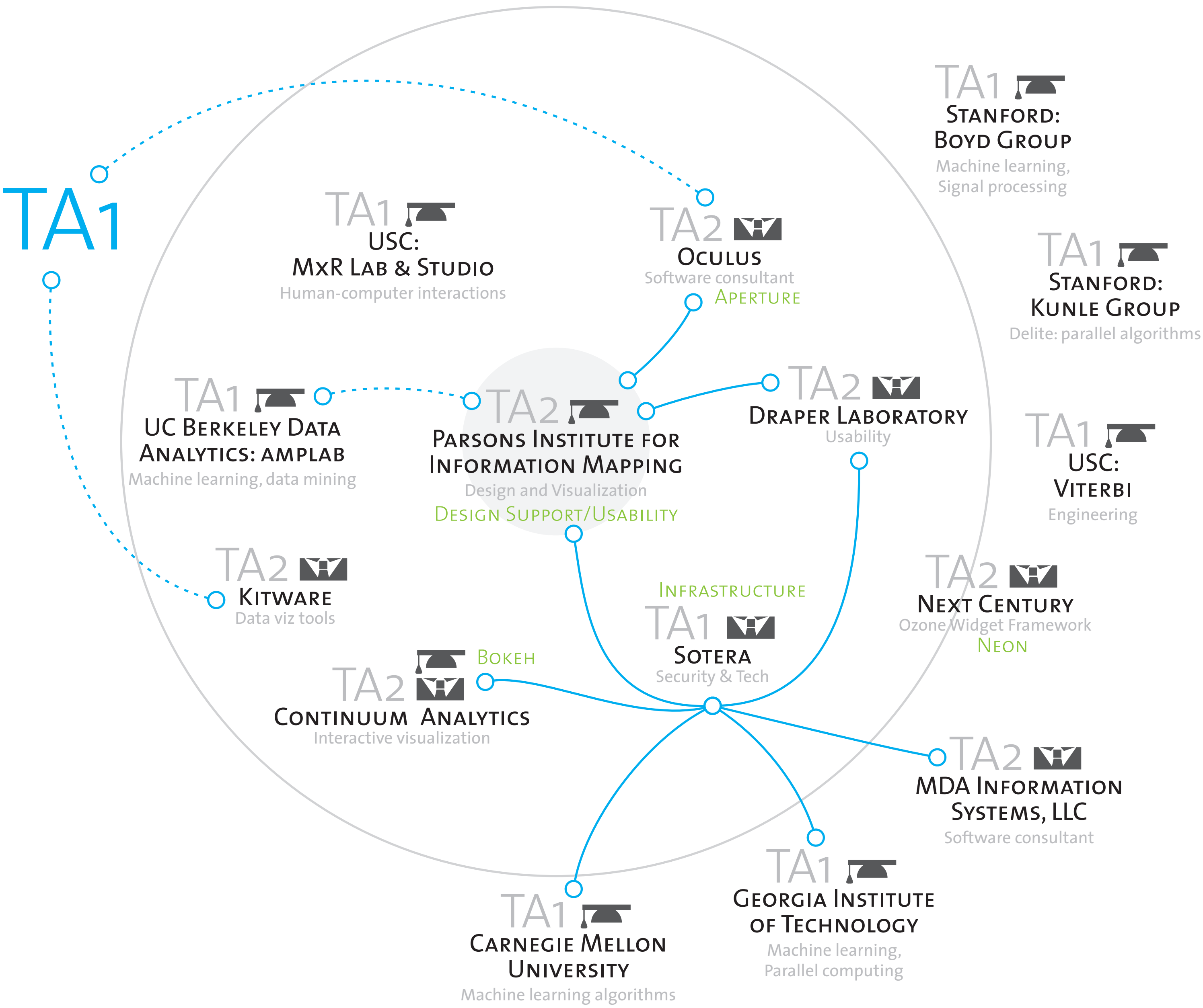
XDATA Components



TA 2 | Parsons Institute for Information Mapping

XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM Parsons Institute for Information Mapping
Bernd Riedel
Christopher Goranson



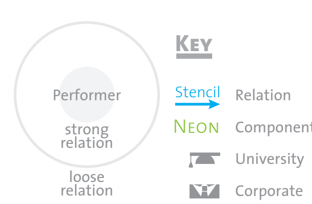
Performer
Parsons Institute for Information Mapping

Domain
University R&D

Expertise
Design and Visualization Best Practices

Project Goal
Research trends, document goals; develop GUI style guides, taxonomy of visual styles; design concepts; best practices for design. Prototyping the XTOOL - concept and application

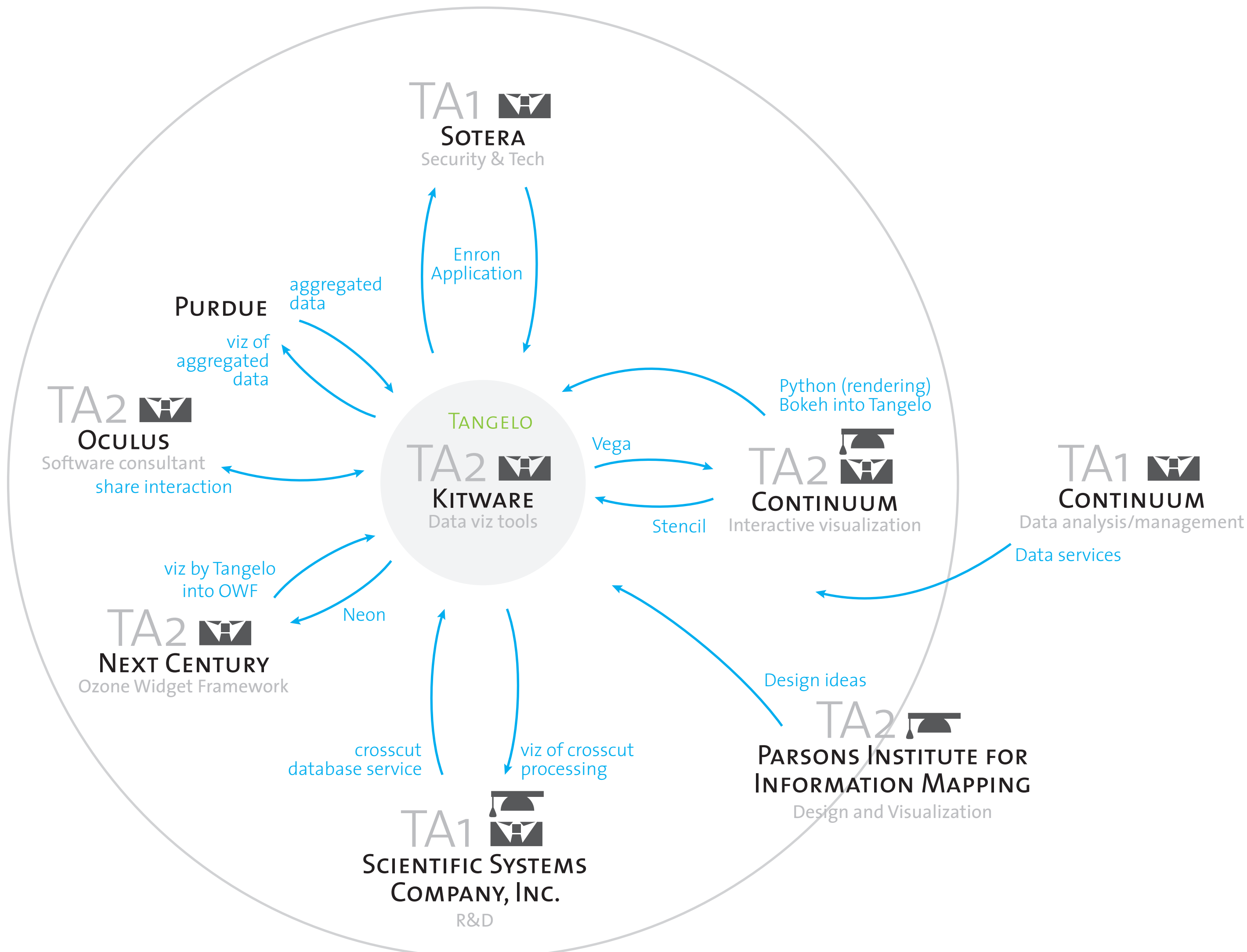
XDATA Components



TA 2 | Kitware

XDATA TA2 Self-Mapping | 2nd TA2 Workshop | Washington, April 18th - 19th

PIIM PARSONS INSTITUTE FOR INFORMATION MAPPING
Bernd Riedel
Christopher Goranson



Performer
Kitware

Domain
Corporate / Software consultant

Expertise
Data visualization tools, visualization design methodology, software development process

Project Goal
Develop the Visualization Design Environment (VDE), an interactive interface that will enable rapid development of visualization solutions with no programming required, using the Vega visualization grammar

XDATA Components
Python web server, Javascript/HTML5 for visualization, and MongoDB