Progress Report

TIC Working Group E
Evolutionary System Architecture

July 2004–July 2005

Group Homepage: http://www.seis.utah.edu/anss/wge/

Prepared for
ANSS National Implementation Committee

September 17, 2004

Outline

1. Working Group Members

- Walter Arabasz, Chair *(Univ. of Utah, NIC & TIC)*
- Ray Buland *(USGS Golden & NEIC)*
- Art Lerner-Lam *(Lamont-Doherty Earth Observatory & IRIS)*
- Phil Maechling *(Univ. of Southern California & SCEC)*
- Tom Murray *(USGS Anchorage, NIC & AVO)*
- David Oppenheimer *(USGS Menlo Park, NIC & CISN)*
- Rick Schult *(Air Force Research Lab, Hanscomb AFB)*
- Tony Shakal *(California Geol. Survey/Strong-Motion Instrumentation Program & CISN)*
- Mitch Withers *(Univ. of Memphis & NIC)*

2. Charge

"This working group is responsible for defining an evolutionary path for transforming existing elements of ANSS into a functional nationwide system—with emphasis on steps that can be taken in the near term (1–3 years), based on realistic ANSS funding projections. The approach should include clarifying key system performance goals, characterizing "where we are now"—both region by region and as a whole—in terms of being able to meet those performance goals, and accounting for geopolitical realities as well as abstract ideals in designing an ANSS system architecture."

3. Foreword

This working group started its work only recently in July 2004. As tasked, it is awaiting reports from Working Groups A, B, and I (and other information relevant to the review of seismic networks) in order to clarify key system performance goals for ANSS and to assess current system capabilities as its starting point for evolutionary system design. The group is off to a good start and has had vigorous discussions about issues and guiding principles for ANSS "system building."

In order to streamline ANSS, some group members want to radically rethink how data processing gets done by components of ANSS. At the same time, the group recognizes the importance of developing a compelling "Road Map for Partnership" for any proposals to significantly change the status quo. ***The ideas described here are preliminary—with lots of committee discussion and work remaining!*** I've intentionally not attached names to the contributions from individual committee members for two reasons. First, so the ideas can be judged on their own merit and, second, because whatever is ultimately recommended will be the consensus of our entire working group. — *W. J. Arabasz*

4. Progress to Date

- Nine-member working group formed by June 30, 2004
- Two conference calls and many e-mail exchanges dealing primarily with identification of key issues, guiding principles for ANSS system design, and models for ANSS system building
- Web site (http://www.seis.utah.edu/anss/wge/), including Internal Web Page (containing contact information and group e-mails, among other things)

5. Quantifying "Realistic ANSS Funding Projections"

Our charge instructs us to define steps for system building in the near term (1–3 years) "based on realistic ANSS funding projections." In order to "guesstimate" future ANSS funding, we used a PERT project-management approach — [optimistic scenario + (4 x most likely scenario) + pessimistic scenario] / 6 — and got feedback from USGS managers in August 2004. The results summarized in the following table have to be tempered by subsequent developments that suggest that even these modest projections are optimistic, and the projected growth in FY07 and FY08 is highly uncertain.

Guesstimate of "Realistic" Funding Projection for ANSS
August 11, 2004

FY	Appropriation (millions of $)	Increase (millions of $)
04	4.40	0.50
05	5.53	1.13
06	6.68	1.15
07	9.22	2.54
08	11.80	2.58

6. Some Key Issues

Assured Response to Large Earthquakes.— One of ANSS's fundamental goals is highly dependable rapid response, especially for damaging earthquakes. What design elements will assure performance/response for any large earthquake in the U.S.? Can regional centers located in seismically active areas realistically be relied upon for effective network performance and response if they're hit by a large earthquake?

Road Map for Partnership.— How do we reconcile state/local ownership, investment in, and/or ongoing support of significant infrastructure for seismic monitoring with the prescriptions of ANSS decision-makers? The motivation for ongoing state/local support should not be undercut. Further, the automatic incorporation of all data into ANSS can't be taken as a given without regard to originating source. Given a natural tension between the idealized design of an ANSS system and the current configuration and funding of seismic networks in the U.S., a "Road Map for Partnership" is needed between federal and state/local elements of ANSS. In particular, **we need to persuade network operators (and their varied sponsors) to move ahead toward a better-designed nationwide system that offers a win-win deal for both individual networks and the system.**

Separation of Routine Processing from Development & Interpretation Functions.—

OFR 02-92 (the original TIC report) argued strongly for this recommendation. There are many pro's and con's. The recommendation has great merit as a basis for creating the underpinnings of a professionally operated ANSS. It also may be a necessary step towards the ANSS goals of "highly dependable rapid response" and "uniform, high quality earthquake products." Network staffers now have to cover so many bases and have to multitask so often that they can't give the latter ANSS goals the dedicated attention they require. When faced with "Zerrissenheit" (torn-to-pieces-hood), maybe this is "forced sanity."

The key to making such a separation work is to establish clear and efficient communications between these functions, so needed changes can be incorporated into the processing. At the least, there needs to be transparency into the centralized processing for the local interpreters, or institutionalized uniformity of distributed processing at local centers—and feedback mechanisms to correct problems.

Resources.— ANSS needs to have dedicated software resources (programmers, etc), and needs to have personnel dedicated to keeping the institutional communications working (continuous training and/or communication between central processing outlets and local interpretation outlets). Both of these are often overlooked.

Centralized vs Distributed System Functions.— OFR 02-92 argued for four modular building blocks (data concentrators, operation centers, data archiving facilities, and outlets), and it recommended that there be one primary operational center per ANSS region. The latter recommendation effectively was disregarded in the subsequent development of the California Integrated Seismic Network (CISN). The CISN experience offers many "lessons learned" for ANSS system building, many good and some cautionary.

Individual Point of View #1. —

- ANSS is a distributed system development effort and the most critical issue will be clear communication between participants. All available techniques must be used to help

address this issue. There is no one solution. We should try many approaches to find some that help.

- Important management issues including "who does what", "who's gets to decide", and "what leverage does management have to motivate groups" must be resolved in order for the technical groups to be successful.
- The [system architecture] group must have a way to prioritize requirements. We will constantly be faced with tradeoffs and we need a way to pick between options. (Fast, Good, or Cheap).
- How can we validate the system works properly while it is operational?
- Do we build or buy aspects of the system, particularly software? Are there any "given" elements that we must weave into the architecture?
- Many groups will be working on the same problems. How can we eliminate duplication of effort within the system?
- The system must recognize the important role of computer personnel within the system and must provide them with appropriate recognition and support.
- Does ANSS management recognize that the software infrastructure may require support equal to the hardware infrastructure?

Individual Point of View #2. —

With regard to the current configuration of seismic monitoring, the word "system" in ANSS is a misnomer. It's more a loose confederation of independent operators who speak to each other in the most limited sense, issue different information on the same earthquake, have various technical capabilities, and archive their information with different levels of completeness, timeliness, and in different formats. I believe the architecture proposed in OFR 02-0092 perpetuates this situation instead of correcting it. . . . Instead, I think that regional networks should function primarily as data collection nodes and that there should be *centralized* reporting, data processing, and data archiving for the entire system. Regional networks should function as backup to the centralized system (i.e., not as primary). I understand the political difficulties that this presents, but from an operational perspective it makes many things simpler and more logical.

Assuming that such a system could be designed, it follows that there actually could be one solution. I maintain that we now have enough experience under our belt to design that solution. It would eliminate poor reporting performance by some (all?) regional networks, simplify software management, and address 7X24 staffing issues.

Individual Point of View #3. —

I agree that regional centers are backups to the national center. But this implies that regional centers are doing the same things for the appropriate region that the national center is doing for the globe. I do not think we should become overly centralized from an operational perspective (which should not be confused with other aspects of the system). . . . By becoming too centralized I'm concerned that we'll lose the local interest in partnering as we inevitably lose interest in meeting local priorities. Yes, its a political minefield that we shouldn't shy away from but it also makes fiscal sense. . . . If you don't have local people routinely analyzing local earthquakes, you will lose your local expertise.

I think we do need to become more solidified technically. Let's consider how a nationally centralized system might work. Unless we put all the knowledge into the software, then the knowledge needs to reside in people. I think the national center will still wind up having regional teams/experts. These are the ones that will know all the gotchas (e.g., where the mines are and what the waveforms look like, which are the best stations, what velocity model to use, which stations have odd responses and converted phases, when to say something just doesn't look right, etc, etc). Does it really matter then where these teams/experts are physically located? Will they benefit more from being with the other processing teams or from being closer to where the interest is most likely to be?

If all we're running are the NSN backbone, urban strong motion networks, and structural instrumentation, then it makes a great deal of sense to do it all centrally with a single backup center. If we want a lower threshold, more precise locations (especially depth) within relatively active zones, and regional and local networks, then I think the centralized processing would have some significant limitations. Of course this is all just a prediction of how it might work and I'm leery of forecasting let alone predicting.

[More Discussion in Section 8 —"Models"]

7. Some Guiding Principles

Explicit Customer Focus.— NIST's nationally recognized Baldrige Criteria for Performance Excellence (http://www.quality.nist.gov/Business_Criteria.htm) provides a systems perspective for managing and guiding any organization such as ANSS toward performance excellence. One of the crucial core values is customer focus. (The customers are not us!) One of our group members describes that the U.S. Air Force has been using the Baldrige approach for five years. He judges that the approach is "one way to enforce common sense" and that it can usefully help organize people. The approach can be helpful to get beyond "political" hurdles to consensus. For our purposes, the Baldrige approach provides relevant guidance not only on how to serve customers but how to design a system.

Endgame.— The endgame is to have an effective system operated with a high degree of professionalism by dedicated well-trained staff (consider the public face of a world-class airline and the behind-the-scenes efficiency of air traffic control and system management).

Certifying ANSS Elements and Modules.— In order to be identified (and funded) as an element or module of ANSS, that element or module should be certified as meeting ANSS capability and performance standards.

System Development.—

- Evaluate the use of a layered, rather than a modular, approach to system development. Layered architectures, with layers isolating changes, have proven to be very successful. Can we develop an architecture in which programs interact with the system at specific layers, rather than through specific modules.
- Due to limited resources, we should focus on making 90% of the customers happy. The last 10% will take an inordinate amount of resources to support. An example in the last 10% might be Early Warning, or Earthquake Prediction. This does not mean the system should work only 90% of the time.

- Address the administration of system through well-defined "roles." Within each regional network all roles must be filled. For each role, associate a name with that role.
- Develop an emphasis on automated testing capability. No system is assumed to work properly without acceptance testing.
- Encourage system and software development through open source codes.
- Introduce test and maintenance schedules to improve reliability.
- Recognize the need for non-data flow infrastructure (communication tools, document exchange, reference manuals, development sandboxes).
- Develop a career path for the computer staff.
- Handling of large events is the most challenging for local networks. Special care should be taken to verify the system (people, software, hardware) is ready to handle large events.

Miscellaneous.—

- Monitoring is not research—and change, in general, is undesirable.
- Long term consistency is crucial.
- Must recognize local mandates.
- Redundancy not duplication.

8. Some Models for ANSS System Building

Model A — Centralized data processing

1. There is only one earthquake. Therefore, there should be only one report about it that reflects all available information.

- Imagine a universal associator that can integrate phase picks from both regional and global networks.
- Imagine software that can choose the correct locator algorithm and velocity model based on the distribution and geospatial extent of reporting stations.
- Imagine a system that constantly updates itself as new information comes in and has an infinite memory such that a new piece of information weeks later can seamlessly update the information.
- Imagine state-of-the-art software applied uniformly to all data in the ANSS. Moment tensors, ShakeMaps, and finite fault estimation for all regions. Locations based on 3-D models. Near-real-time waveform cross-correlation and double-difference locations. Dare I say it....Early Warning?

2. Bandwidth is infinite, or soon will be. Let's take advantage of it. For example, the NCSN currently has ~1000 channels at 100 sps, which is approximately 3.6 Mbps. The TIC document Appendix B estimates an aggregate bandwidth of 17Mbps for the

full-blown ANSS system. Child's play for the Internet, modern networking hardware, and computers.

- Imagine a centralized system that is acquiring continuous waveform data from regional collection nodes

- Imagine that this system would properly save all the SP/SM waveforms from an event in a centralized DBMS as well as continuously archive broadband data.

- Imagine that professional data managers/system operators would monitor the flow 7X24 and ensure necessary Q/C takes place. (Contrast this model with the OFR 02-92 model where each of the 7 regional centers try to do this).

- Nervous about the Internet? Dedicated circuits (with failover to the Internet) would probably cost on the order of one or two highly paid seismologists per year.

3. The user community deserves a single portal for information. We have finessed this problem with QDDS/QDM for the recenteqs and ANSS/CNSS catalog, but it's a kludge (come on...admit it). Users now have 6 places to go for waveforms (the IRIS DMC, NCEDC, SCECDC, CISN Engineering DC, NSMP DC, and COSMOS DC). I have no idea where to go to get uniform phase information for the entire nation. Catalogs of mechanisms, ShakeMaps, processed strong motion, etc are scattered all over the place, if available.

- Imagine there is but one place where users go for information. Perhaps behind the scenes the something like NETDC is at work where the system knows about other datacenters and can farm out waveform requests accordingly.

- Imagine that all data are uniformly processed in a timely fashion, that the data/methodology are well documented, that there are similar levels of information for all regions of the ANSS

4. People are expensive. Digital communications makes it possible to have data analysts work in low-cost areas. Picking seismograms is not rocket science, particularly if the analysts are trained and the algorithms guide the analyst to the correct interpretation of phases.

- While run-of-the-mill locations and data review would be done by dedicated staff in low-cost areas (but not overseas!!!), professional seismologists in the regions could also review more complex events that call for finite-fault estimation, etc.

5. 7X24 rapid response is a "must" for the ANSS. Why staff 7 centers round the clock when digital communications makes it possible to do this centrally at lower cost?

6. There should be 2 redundant, centralized systems in separate locations of the country to ensure operation if one becomes non-operational.

- Imagine that these 2 systems replicate their real-time and archival databases in near real-time

7. Regional centers can generate earthquake reports for their region in the event that the 2 national centers are not functional. They would not back up global reporting, however.

8. Regional centers would have weeks of waveform storage in case communications to the 2 national centers is severe.

- Imagine a really big quake in a region.

9. Science is local, but processing is national.

- When quake occurs in a region, the regional experts will be expected to respond to press/OES inquiries. This is exactly what happens now (except for the general statements that NEIC releases about regional events). The press knows to seek out the experts.

- Regional seismological research would be expected to feed up into national capabilities. For example, when a new and improved 3-D model is developed for a region (or continent), the national system could integrate it. When some smart dudes develop the successor to the double difference algorithm (triple difference?), a national system can integrate it for all regions.

Model B — Outsourcing of data processing

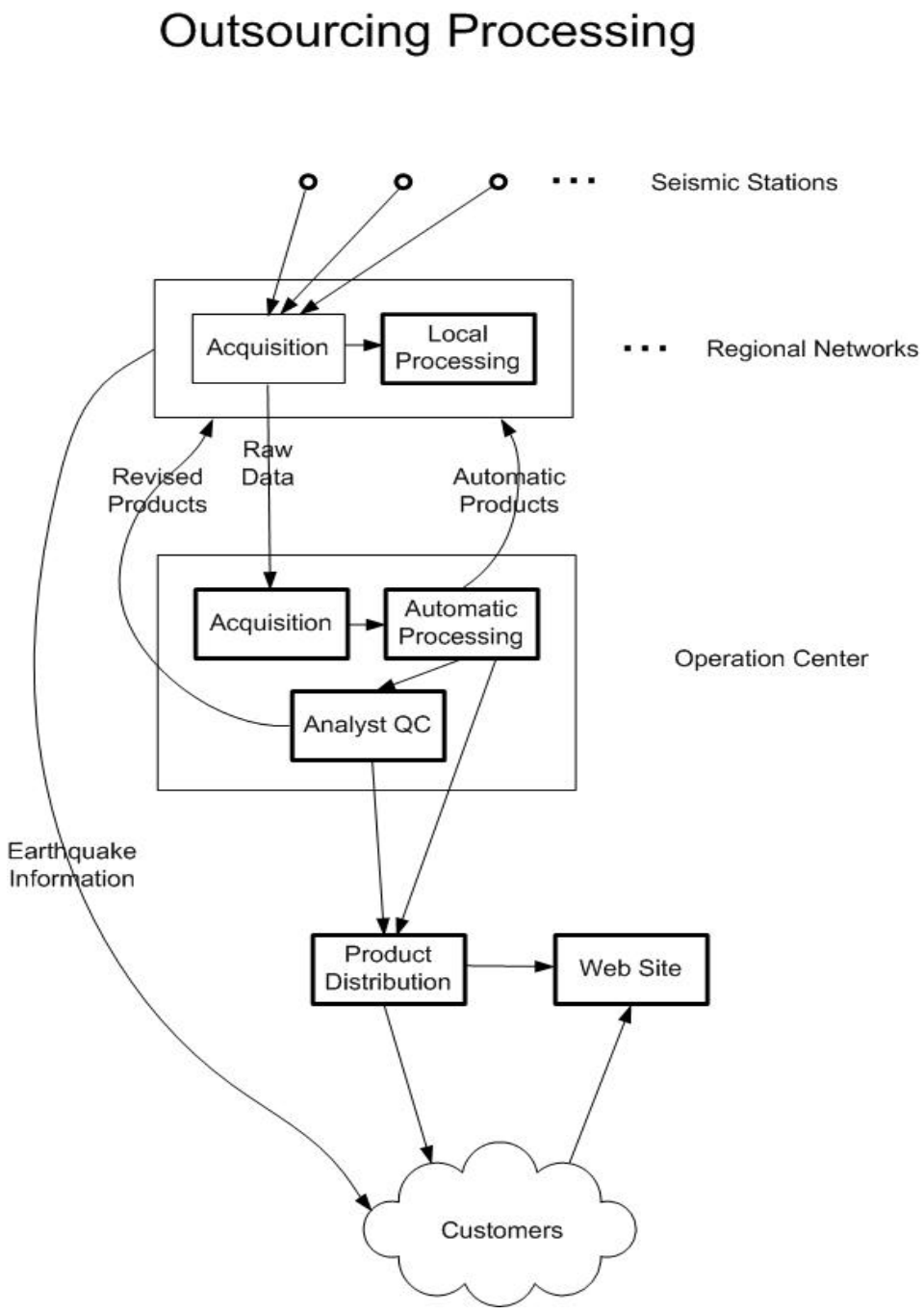
How to take advantage of economies of scale without losing the expertise of key personnel?

I have outlined one way that this might work in the block diagram below. Because of the parallel with business, I've called this outsourcing. I think the business analogy is useful. Large companies are often organized into departments that provide unique products or services (core businesses) and support groups. Often the support products or services were developed in-house when there was little alternative, but can now be done better and cheaper by companies that specialize in the relevant commodity. By outsourcing the support product or service, the large company can focus on their core business to try to improve their product and lower cost at the same time.

Of course, this is exactly where the ANSS finds itself in the current evolutionary environment: with a plan to improve products and services, but a pressing need to lower the cost of doing so. The diagram shows one model for accomplishing these goals. The idea is for the existing regional networks to act as ANSS data concentrators and pass the raw (waveform) data on to a central operation center for processing and QC (analyst review). The automatic and revised products (e.g., location, magnitude, moment tensor, shake map, etc.) produced by the operation center are passed to a product distribution mechanism (i.e., AFIDS) and also back to the regional networks.

The regional networks then interpret the products for emergency response personnel, local officials, the media, etc. In other words the local experts perform the key role of converting earthquake data and products into earthquake information. Notice, that there is also no reason why the regional network centers can't crib a copy of the waveform data as it passes through to the operation center. This can be used to support special projects, research, and education locally.

In this model, interpreting the data and products for the public is the core business of the ANSS. This is preserved by retaining the local expertise where it currently exists. The dog work of automatic processing and the QC of products is outsourced to the operation center service bureau. As with any outsourcing, it is important that the service bureau is responsive to the needs of the community. This would be accomplished by building on the experience of the regional network operators (e.g., local models, corrections, and known source regions). It will be important that the community continues to oversee this operation to ensure that it meets ANSS needs.



Model C — Performance-based assessment and funding

For completeness, and what it's worth, I'd like to describe an ANSS architectural model that was discussed some time ago during earlier technical discussion. I'm not claiming this is my new idea, but it does seem to be a model we haven't considered. I think this is a type of a "performance-based" system.

In this model, the system analogy is the electrical grid. In California, if I produce excess electricity, and I can deliver it at 120V/60HZ, SoCal Edison will buy it from me. Edison doesn't care how I generate it as long as I conform to their interface.

Applying this to ANSS, we (ANSS) specify the quality of data that we will accept, and establish a value for a data stream ($/hr or something) for streams that meet our level of quality. Higher quality streams (broadband) should be more valuable. There probably would be a need for "zoning", with rate adjustments for "hard to reach", or "under represented" zones.

Local systems are free to operate as they wish. As long as they provide us with quality data, ANSS doesn't care how they operate. The ANSS funding to networks is based on quality and quantity. This relives the ANSS architecture from worrying about how all the local networks work. We define data format standards, quality standards, data transfer mechanisms, and networking topology, and wait for networks to plug in with high quality data feeds.

Does this model offer anything new? I think this may actually be a variation on [Model A], which confronts the reporting responsibility issues more directly. This description addresses the funding and division of responsibility between regional and ANSS more directly. This model sharply limits how involved ANSS [should] get into regional network operations. Both models seem to imply that there is a large ANSS real-time earthquake monitoring center, and a data center that reports on [all small seismic events] around the country.

9. Next Steps

- Meeting of 5–6 members planned for Sept. 27, 2004, in St. Louis
- First face-to-face meeting of full group planned for mid-November 2004, piggybacked onto ANSS-NSC meeting in San Francisco
- If warranted, arrange informal meeting of working group members during Fall AGU meeting in San Francisco (Dec. 13–17, 2004)
- During next two months, use results of other TIC working groups to (1) clarify key system performance goals important for system design and (2) assess “where we are now” in terms of being able to meet those goals
- By end of calendar year, reach consensus on recommendations for either adopting or modifying what was proposed in OFR 02-92 for ANSS architecture and interconnection
- Set target in Spring 2005 for (1) completing design of a realistically achievable ANSS system and (2) prioritized steps during next 1–3 years to approach that outcome, to the greatest extent possible