

*Speech to BASIS Government User's Conference at the University of Maryland –to speak to our language needs and some of our recent successes partnering with industry to solve them, May 2008*

Good morning and thank you for inviting me to your conference. This summit addresses a critical need – linguistic intelligence which allows us to understand the people we interact with around the globe.

Having the right technological tools is absolutely essential to CIA's mission. In most cases it does not make sense and would not be effective for CIA to design and build a solution itself. The private sector often leads innovation and CIA must tap it for new tools, adapt them as necessary for our unique needs and transfer them quickly into our operations. At CIA, we rely on scientific and technological innovation to give us the edge we need in both collection and analysis.

CIA is a worldwide organization. We must operate wherever the mission takes us, often in ways that leaves no footprint behind to...

- Move people, data, and equipment around the globe seamlessly
- Connect people across geographic, cultural and language barriers
- Manage and search large data sets
- Provide situational awareness to support decision making for operations and policymakers.

Meeting all these needs requires that we improve our language capabilities. In our case, the human solution is overwhelmed. Disciplined, regular investment in technology is our way forward. And in an ops oriented organization – language technology is one of the very few areas in which we've been able to sustain long term investments.

During the last several years at CIA, we've come to rely on In-Q-Tel as the Agency's window to the breadth and depth of America's entrepreneurial culture. In-Q-Tel fosters technology development that has specific value for the intelligence mission.

Within its broad portfolio, In-Q-Tel has a significant block of language technology investments. Language processing was a wall we had been beating on for years. A plethora of research had been conducted and many technologies had reached the stage of being capable of moving to operations. But they hadn't been able to make the leap from the bench to the field. Over the past few years we've made that leap.

For example, the investment that In-Q-Tel has made in technologies that automatically screen and add structure to raw data greatly enhances our ability to recognize actionable intelligence in any and all of the sources from which we obtain data.

Our host today, BASIS Technology, is an example of the contribution of the In-Q-Tel portfolio. BASIS has done groundbreaking development in automatic vocalization, creating an algorithm that allows us to convert names in foreign alphabets that do not explicitly represent short vowels in print. Without BASIS technology, the converted letters of the name Mohammed would be 'mhmd' which is neither easy to pronounce nor to understand to search for.

Names of people and places are among the most important pieces of information that an intelligence organization can have. BASIS' vocalization and transliteration algorithms made millions of names and their relationships available to analysts in days rather than months. Overall, this technology has saved hundreds of man-years of labor and millions of dollars, and it has produced strategic information resources we would not otherwise have.

Language Weaver is another company we've worked with to develop a technology that has successfully made the leap from bench to field. The Language Weaver implementation of statistically based machine translation technology allows useful MT systems to be created in less than one year, with translation rates in excess of 65,000 words per minute. Our analysts can now access in minutes and hours what would have taken months and years of tedious manual translations by linguist. The time saved is vital to our ability to take action.

Language Weaver also allows intelligence community users to automatically customize Language Weaver's 'out of the box' technology to perform better on our kinds translations. A Language Weaver module lets us recycle our past human translations of classified documents as training data.

So to summarize our need for language technology

- We need to be able to scan raw data of interest and make it immediately available to every analyst, regardless of the analyst's language ability.
- We have to make the connections between shreds of data in any language.
- We need to accept data and be able to determine on the spot if we have seen it before, and if it is genuine.

This is a tall order and requires that every option be considered. Every one of you attending today can help. Thank you for your continued support to our community.