

## SensorNIS2011 – Porter notes

- discussion Session 1
  - Need funding to coordinate - not just data standards
    - definitions are still an issue: your calcium may not equal my calcium
    - did have a lot of work on coordination - often separate
      - everyone was leading
  - many technological advances that make data sharing easier
    - barrier is much more one of organizational structures - multiple mandates for the federal agencies
    - eg Data.gov
      - CUASHI went to USGS, EPA etc. to get them to share standard
    - Universities can play role
      - to stabilize networks need to interpret the data - show value
    - at USGS have been working backwards from policy goals to measurements
      - did you get far enough to do 2-way flow
      - need to create bridge between scientists and information managers
        - but currently are 22 enclaves
      - need to link long term monitoring to research
  - for NE may want to look at what policy or management question that sensor network would allow us to make easier decisions
    - did that at DOI, but never had enough interdisciplinary data to show value
    - Yukon Basin may be good example
      - lots of scaling issues
      - logistical issues
      - 3 years of methods development
        - based on science questions
      - policy questions - maintaining villages as environment changes
        - also used to guide science
          - got method for mapping permafrost over wide areas
          - tracking carbon down rivers
          - permafrost carbon could be run-away train
        - also wanted to leave legacy program as model for other systems
  - Bridging cultures not just agencies
  - NEON may be foundation for national network
    - USGS planning to do additional monitoring
  - want legacy of key locations
  - long time frame develop
- issue may be more creating good quality controlled data
  - will talk more about barriers this afternoon
  - what are barriers
    - institutions
    - technology

- money
- expertise
- funding model- need new idea, research
  - but we need products that work
- then drill down to nuts and bolts
- I think expertise and field deployment
  - lots of sensors work well in lab, but not necessarily in the field

are also opportunities

- can crosswalk between standards
- need some additional glue

some data are no-brainers

- high frequency water quality data

barriers - complex environment of vendors and standards

- lots of site-specific approaches
  - tool sharing consortium would help
- data models and services layered on them
- but complexity is just moving up the food chain
  - some tools work for 20% of data
  - another tool works for another 20%
  - some data fits multiple models
  - eg. time series vs gridded
- may need to look at community level standards
  - soften barriers between stovepipes
  - bridge time series and gridded data
  - make tools for translation that do behind the scene
  - makes me sceptical about semantic approaches - need to define things in ways scientists want them
- need to focus on interfaces between different system components

data quality and assurance

- sensors come and sensors go, sensors drift
  - most programs don't do a great job on calibration
    - exceptions: NOAA CRN
  - issues - subtle changes in gain
    - NEON will be focusing on this
    - hard to do in short-term projects
      - not paid for by grants
- some sensors are so new, we are not sure what they are telling us
  - e.g., aquatic sensors
  - I don't feel comfortable putting it out - don't know what it means
    - data labeling may help
    - provisional
- can QA against regional models
  - identify breaks in records
    - found lots of times instrument or location was changed
  - need to track calibration
    - don't have good tools for tracking

NEON has tried to address - and we have funding to do it!  
astrophysist view - we struggled with the same issues

- work on X-ray observatory
- about 30% of costs are related to QA
  - have a whole calibration team
  - can reprocess entire data stream
  - documentation of plan and details of calibration
- do a lot of cross calibration
- customers
  - Xray and radio folks want integrated data
  - but Xray folks won't be expert on radio and vis versa
  - need filter to get it out to the rest of the world
- need to know all the steps, calibrations etc. 5 years from now

Need cultural change

- NEON ahead of LTER which is ahead of other agencies on IM
  - USFS culture does not spend money on IM
  - barrier of finances
    - better data vs more data

also need way of testing data for homogeneity when detailed calibration record does not exist

- uncertainty analysis

users will use data, however bad....

- need to meet
  - documentation needed
  - requirements
  - operational
  - testing to see if meet requirements
  - users guide

cultural shift

- need to see what will work for graduate students - learning by doing
- if people are expecting to use data off web, then proper documentation becomes more important

differences between data from well funded teams - and not!

- flux data- no overall quality control
  - data in same form is shared
  - but students have no idea that some sites have problems
    - leads to spurious results
- software package for doing QA could be used by others
- e.g. R open source
  - developed jointly
  - not tied to individual companies
  - still does some changing - some parts not as mature as other parts

need to make it politically relevant..... and confident in quality of delivery from network

- chicken and egg problem

- Boose - questions
  - what about storing code vs data versions
- afternoon discussion
  - how to balance new science, new sensors , more sensors, processing effort
  - maximize analysis
  - see Corrina's PPT notes
- Sensor Tech. & Software
  - success stories
    - USGS
    - weather service
    - challenge often hard to find streaming data
      - usually raw
      - use at your own risk
      - need to be confident in data
      - had to go to multiple sites
      - data turbine
      - levels might help
    - huntington forest
      - stage height, climate
      - not sure what LEVEL it is
    - discussion of levels
    - mandate for making data public and available
      - priority
    - QA QC should not just be on the data
      - hardware, software also included
    - PI driven projects vs large projects
      - PI product-driven
      - vs customer driven (NEON)
      - no naked data, every step documented
  - focus on steps
    - collection
      - NTL has buoys
        - lots of parameters
        - issues with freezing
        - often use off the shelf
      - identifying the right sensors
        - level of precision
        - choices are there
        - we all have our own sensor suite....
    - loggers
    - who makes decisions on the types of sensors to use or to collect data
      - often recommendations move upwards
      - used blog for techs to chatter
      - budgets are also a consideration
    - should be science driven

- don't trust vendor specs - in house calibration
  - but does that happen or is it vendor driven
  - often a variety of needs and capabilities
  - mostly COTS
- needs
  - should have mandate for data sharing
  - low cost sensors that we can distribute in space
    - motes
      - what do you sacrifice to get low cost?
  - gap between motes and traditional loggers in terms of spacing between stations and cost
- streaming
  - information sharing is key - static and interactive
  - what are barriers to doing this?
    - depends on domain
      - marine has different issues
  - sometimes permitting issues
    - no towers
    - satellite can be expensive
  - successes - David Hughes Wireless Biology Project
  - best practices outline would help
    - need something on site to buffer
    - radio
      - also wired
        - NB485
    - bandwidth needs
    - range
    - terrain
    - commercial service availability
    - costs
    - serial vs multiplex (ethernet)
    - power
      - line power (ideal)
      - wind power
      - solar power
      - fuel cells
      - microwave power
      - issue - hard to calculate
        - issues - inefficiency
  - can do nested hybrid
- Data Access
  - unfettered access to raw data?
    - can share as raw
    - need to document events in data
      - cleaning wires etc.
    - issue of priorities - would be a huge documentation task

- can be in lots of different formats
    - EPA IMPACT wanted to do real time streaming
      - EPA nixed
    - real time data would be useful
  - need to identify users and their needs
  - would like to develop standards for levels of data quality
    - NEON is similar to NASA
  - what is controlled by grants, contracts, deliverables?
    - users
    - stakeholders
    - sometimes mission-driven
  - mechanisms for customer feedback?
    - issues of privacy and intrusion
    - email of IM is available - field questions
    - finding papers using data
      - often papers don't say which data they used
    - fund source is listened to
  - real-time - many users are outreach users
    - station manager uses
    - quick-look
- opportunities
  - ways to share information on instruments
    - some are well understood
    - others are newer and less well characterized
    -
  - what kind of forum would work?
    - individual contacts
    - EFR techs mailing list
  - metadata often lacks details on instrumentation
    - and solutions
  - want to build community of practice
    - meetings
    - VTCs
    - listserv
      - needs to have archive
    - WIKI
    - effective, efficient and non bureaucratic
      - term community is very important
    - FaceBook
    - CampbellSci has list that can be useful
    - geochem listserv
      - big community
      - lots of topics NOT of interest
      - they have someone responsible for maintaining - moderating
  - issue -often things die on the vine

- needs to be critical mass
    - funding might help develop social networking
    - surveys of what is being used would help
    -
  - stratify discussions by research topics and user communities
    - within group some sensors are well known
    - sensors are group specific
    - communication within meetings
    - blogs can also help with sensor selection
    - grad students can help research
  - seek funding for developing social networking tools to share experiences
  - develop best practices for wireless
  - best practices on how to share data
    - granularity
      - single large files
      - lots of annual or station files
    - query based
      - how to keep from bogging down
        - database design issues
      - best practices for database development
    - can do both
    - one form doesn't fit all.....
- QAQC working group 3
  - real time data
    - range checks
    - missing value codes
    - minimum could be nothing
  - issue: is it responsible to release unapproved data?
    - end users
      - technical and logistical use
    - can watermark graphs and not release data
    - people can pick numbers off scans
    - but we can't be responsible for how people might misuse data
      - our responsibilities are to tell them what we did to check
      - they are responsible for their use
  - good practice
    - varies depending on scale/number of measurements
    - some common ones
      - range/min/max
        - definition is subjective
        - can get from sensor manual
          - that deals with impossible, not unlikely
    - easy to use tools may help
  - levels?
    - work in context of individual systems - but do not necessarily apply across systems

- usually
    - level 0 raw
    - level 1 calibrated
    - level 2 gap filled
  - can put things in metadata detailing
  - lots of domain choices
- running averages - lengths
  - specific to datasets
  - can calibrate on existing data
- redundant sensors would help
  - can use nearby stations
  - multiple loggers also help
  - 3 is ideal number
    - increases costs
    - don't all need to be of equal quality/resolution
  - also increases reliability
  - multiple sensors of different types
- seven tests are pretty good but don't apply to everything
  - how conservative should you be?
    - may cause more harm than good if strip out interesting points
  - ultimately you are responsible for your choices
- are methods for doing GENERIC QA - but no easy to use tools
- tools?
  - lots of candidates
  - Matlab/GCE toolbox
  - Could have some R libraries
  - statistical languages/programming languages
    - code can be put in metadata
  - kepler
    - can share agents
  - web service-based approaches
  - proprietary software
    - Campbell RTMC VistaVision
      - and other graphical software
    - LabView
  - good to have toolkit - that can be customized for data
- education and training
  - need training workshops
- gap filling
  - is it nobler to gap fill?
    - very useful but also dangerous
    - need to make clear what was done
  - use of nearby stations
  - lots of other potential models
  - from internal temp sensor could regress to main
    - also weir to wier



- snow models
      - sometimes with boundary techniques
    - backfilling may be a SCIENCE problem - not a IM question
    - Borer ESA Bulletin paper
    - what would you want from a gap filling model
      - lots of tools - lots of approaches
      - if anything too many tools
    - documentation critical
      - should report level of uncertainty
        - can create "fake" gaps to test fit
        - also regression reports
  - clearinghouse for tools and approaches would help
  - will we ever be able to get away from provisional data?
    - even NEON will use human eyes as last test
    - timing depends on availability of human
    - maximizing automation can help
    - automated QC can AID human - guide human vision
    - human won't work for 20,000 datasets
  - would be good to see some tools widely used.....
    - test them on exemplar datasets
      - take good data and corrupt it
      - would also be useful for testing
    - seven criteria are a good start
    - could be used with educational material in training using different tools
    - also exemplar data documentation
  - Action items
    - what should go in best practices document?
    - QA procedures applied must be documented -esp. if NONE
    - Tools - people are eager to use
      - I'd like to see some comparisons and then workshops to select
      - getting a list of on pager on tools
        - can do survey once you have short list
        - articles in DataBIts
      - example analyses in different tools
    - gap filling tools
      - proof is in pudding - test exemplar datasets using different methods
        - what is criteria for which is best
        - want UNCERTAINTY - confidence intervals
        - Mean square error
        - could be a publication comparing methods
      - ameriflux network might be resource - they have done alot
        - they use "gold standard" dataset
        - they published in agricultural meteorology
- Working Group Reports
  - NERC web site
    - educational materials

- middleware
- link NE sites - common data display
- LTER
  - one page descriptions - knowledgebase
  - coding schemes for data flags
  - databits articles for 2012 ASM meeting
  - versioning model issues for streaming data
  - Campbell Paper
  - sensor management system - tracking sensors and history
- Software & Tech Outlook
  - Martin
    - success stories
      - USGS, Weather Service
        - Agency level efforts - mandated and funded
        - not so for PI systems
          - limited funding, need to coordinate collections
      - some individual sites
      - Example David Hughes wireless biology
    - needs
      - COTS is primary solution
        - mostly challenges seemed to be met....
      - mechanism for SHARING information
        - sharing info on customized solutions
        - blogs, listserves etc.
          - require work to keep working
          - needs to be source for EXPERTS as well so they keep communicating
      - best practices would help
      - static documents
      - ongoing communication
  - Campbell
    - Success stories
      - homegrown programs
      - Campbell Sci solutions
        - most sites
      - Matlab
    - common themes
      - use different data loggers
        - Campbell
        - Hobo
        - interest in NEON GRAPE data logger
      - lots of neat software tools
        - outbox
        - flot graphing tools
    - approaches

- top down vs bottom up
  - top down
    - faster
    - more uniform
    - less flexible
  - we are mostly bottom up
  - adopting new solutions - how to make decisions
    - reluctance to invest in things that are not mature
- greatest needs
  - programming support
    - undergrads?
      - not all stable well documented software
  - training workshops
  - Webinars?
  - ways to share experience with tools and software
- action items
  - useful tools
    - perhaps Kepler
  - standardization
    - light handed, top down veneer
- Alene
  - action item
    - way of communicating collective knowledge
      - robust against staff turnover
    - best practices manual
      - format
        - online guide
        - divided into sections
        - links to external resources
        - page manager assigned to update it
        - associated blog/comments used to drive page revisions
    - subsections
      - Managing sensors
        - sensor specific information
        - power sources
        - sensor audits to assess quality control
          - designated staff person
          - sensors that could be temporarily co-deployed for comparisons
      - tracability of a particular sensor
        - GRAPE
        - android phone
        - bar codes

- network level desing
  - streaming
    - review of software packages
      - commercial vs open source
      - custom built
    - telemetry and communcations options
      - tools for design "landscape toolbox"
      - tools for predicting radio pattern
    - archiving
      - citable database
        - versioning issues
        - snapshots
        - annual editions
    - access
      - how to identify and access users
      - ID untapped user groups
- QA/QC discussions
  - Sheldon
    - recommendations
      - NEON streaming data checks
        - complete but subject to interpretation
        - some dependency on sensors and phenomena
          - eg. Range
            - of sensor
            - of reasonable values
            - seasonally based?
      - still substantial complexity - dependence on characteristics of sensor
      - need knowledgebase
      - tracking how long since last calibration
        - sensor management system
      - need to be clear what PURPOSE is of quality checks
        - and how it will feed into data cycle
      - MUST
        - outside range of sensor
        - timing consistencies
      - SHOULD
        - depends on sensors
        - knowledge-base
        - need expert advice

- only missing values (e.g., logger codes) should ever be REMOVED from the data stream
  - Gap filling
    - controversial
    - approach can have major impacts on analyses
    - often needed for SUMMARY data
    - end users want clean, ready to analyze datasets
    - may be done later in the datacycle with expert attention
  - qualifiers
    - many vocabularies available from federal systems
      - impractical to crosswalk
    - good to have a rich set of qualifiers
      - e.g., sensor higher than yesterday
      - for reports
    - but a SIMPLE set of flags for users
    - somethings may be better than qualifers
      - e.g. sensor ID column in data
  - documentation
    - methods, thresholds, assumptions
    - MUST detail gap filling and flag modeled values
  - Action items
    - best practices - crouced sourcing
    - tiers of recommended flagging criteria for particular sensors
- Laney
  - basic QC - Lev0 to level 1
    - data ranges - staged by time of year from historical data
      - need to be careful of extreme events
    - removing vs flagging bad data
      - most prefer just to flag, but not remove
    - permanent vs temporary flagging
      - e.g., tag questionable for further examination
    - numeric vs character flags
    - mature vs new types of measurements
    - need descriptive flags - if data are bad describes WHY the data are bad
  - Variance
    - increased variance can be used as indicator of sensor degrading
    - need to look at windows of time
    - standards of frequency of observations
    - standards for detecting outliers
    - cross-site standards difficult ot establish
  - QC at different points in processing
    - level 0

- raw
  - level 0.5
    - unit conversions etc.
  - level 1
    - QC and QC flags
  - level 2
    - longer time series may prompt further explanation
- gap filling
  - should missing time stamps be infilled
    - if so would you then flag the blank rows
    - or just flag start and end of gap
  - multiple algorithms
  - trust issues
- data qualifiers
  - different at level of processing
  - subjective flags vs. quantifiable
    - good vs passed rangecheck
    - additional flags for if there is additional information
    - should the data be analyzed flag
  - gap filling, drift detection
- data documentation
  - Key - tell how collected and processed
  - metadata requirements
  - maintenance log
  - configuration logs
  - links to other files related to data
- Dereszynski
  - minimum level of QC
    - can be NONE - if we document it
    - what is needed for stakeholders
    - MUST document what has been done (or not)
  - good practices
    - based on NEON list
    - range checks
      - even this simple is challenging
    - running averages
      - specific to site, sensor type
    - sensor redundancy
      - ideally have replication of 3!
      - can use some lower resolution sensors as proxies
      - correlations between adjacent stations
    - help establish confidence in data
  - tools
    - GCE toolbox
    - Kepler
    - R

- web services
      - server-side QC
    - require training and evaluation
      - annotated examples, shared experiences
      - webinars
  - action items
    - series of 1 pagers on tool experiences
      - mix of expert and novice users
      - specific methods
      - performance
        - how well did it work
    - survey of reports in Databits,
  - Gap filling
    - often a science question
    - lots of potential methods
    - emphasis on documentation
    - want clearinghouse for tools and approaches
    - action items
      - create test sets to evaluate gap filling
        - inject known modes of sensor failure
      - criteria for evaluating effectiveness of gap filling
        - uncertainty estimates
      - performance metrics
      - publication comparing methods across exemplar datasets
    - Will we ever get away from provisional data
      - human likely to be part of process
      - focus tools on focusing human attention
- discussion of working group reports
  - need for knowledge sharing
    - can produce guidelines for how community moves forward
    - minimum standards
    - a white paper won't cut it!!!!
      - need creative knowledge sharing
    - key that guides design process
  - best practices document
- subgroups
  - creating knowledgebase web site
  - identifying some common data levels
  - sensor management systems
    - what needs to be tracked
    - how
  - products of the meeting
    - paper outline
    - best practice document
  - discussion

- one group (Alena) had detailed info. on best practices - we can use their outline
- Sensor management
  - goals
    - add to metadata for data interpretation
      - sensor relocation
      - track calibration events
    - needs to be done very efficiently
      - bar codes
    - history of the sensor
    - use to establish calibration
      - nagging system
      - preventive maintenance
    - flagging data failures in data
  - track
    - sensor swaps
    - calibration
    - sensor events
      - wasp nest
      -
    - some will require free text
    - others are events that can be a checkbox
    - tie into photos
      - photos of entire station
    - deployment related things - affect entire station
  - + - mobile phones with API can do app.
    - GPS, compass, camera
    - read bar codes
      - can launch activity
    - Mikhail can do
    - the hardest part is easy to use interface and database it connects to
      - perhaps web services interface
  - sensorML standard covers station, deployment, sensor, parameter
    - can be really complex
    - a subset might be a better way to start
  - could use simple data, object, action
    - with table of ids
    - good to dump to from field ap
    - dovetails well into QAQC
  - best practices
    - tie into provisioning of sensors
    - capture sensor documentation when it comes in
    - good reference tool for building metadata
  - lookup table for standard implementation parameters
    - sensor characteristics
  - could be a centralized system - or a standards-based system



- wiring diagram practices
  - labeling
  - photos
  - diagrams
- action items
  - closer look at SensorML
    - SWE comes with suite of tools and services
      - sensor web enablement
      - also transducerML
    - who is using it?
      - CSRIO in australia
        - also doing marineML
    - id existing utilities and tools
    - obvious choice
  - development of tools to feed off of sensorML
    - may want to develop relational database that feeds off of sensorML
    - notification services etc.
  - funding?
    - RCN -type proposal
    - supplements
    - LTER product-oriented working groups
      - bring in some outside folks
      - or release time
  - Android application development
    - Android app is easy
      - read ID - bar code
      - action
        - sensor added
        - sensor removed
        - sensor calibrated
        - sensor cleaning
          - need time range for this
        - other - text
      - add photo
      - way of recording time span where work is done.....
    - big issue is linking to system
      - a standard would make it easier
      - perhaps standard web service
    - where do you send it....
    - similar to things used in utility industry
    - could possibly add RFID scanner if desired
      - adds cost
      - possibly serial to bluetooth
      - use QR code
  - who
    - Wade and Derek will research sensorML

- set up email list
  - many hands make light work
- Wrapup reports
  - web site of best practices
    - flow - what to think about at each step
  - Campbell Paper
    - reviewed outline
    - areas for future development
    - need good examples of problems with datasets
    - have clear vision of where we are going
    - John Campbell will coordinate
  - data levels
    - agreed on levels 0, 1a, 1b etc.
    - provisional
    - level 2 gap filled
    - will write up for QA report
  - sensor management
    - see notes above
- closing remarks
  - thanks for coming!
  - very productive and fun
  - thanks to Hubbard Brook Staff!
  - esp. Krista
  - organizers - thanks
  - people did a great job paying attention