

Summary of breakout groups as well as general discussion on Data Endorsement for GEOSHARE (see below for detailed comments from each breakout group)

Q1: What is the value of GEOSHARE-endorsed geospatial data products for agriculture, food security and the environment? Who will benefit from such endorsement? Are there downsides from such endorsement?

Different participants had differing perceptions of what is meant by ‘data endorsement’ and the value of such endorsement varies by user community. Private sector stakeholders, seeking to make long run investments in areas where public regulation plays a role, must have a clear picture of the baseline from which they are working. Public sector policy makers need good data in order to make properly-informed decisions. All of these groups turn to the scientific community for guidance regarding data quality and validity. For the scientific community, replicability, peer review publication and open access are key.

At the workshop, Navin Ramankutty emphasized that different datasets are ‘fit’ for different purposes. He produced three global land use datasets – each one giving precedence to different source data. He noted that some users will need to benchmark to FAO data, for example, while others will want to match national census data. Related to this data precedence concept, there was also interest in the AgMIP approach to offering differing levels of endorsement: silver, gold, platinum, etc.

Q2: Different datasets and workflows may be suited for different purposes: Which of these is most pressing for GEOSHARE to achieve, and how might this be accomplished?

Participants noted that the identification of key questions linked with primary scientific and other stakeholder communities will be critical for establishing the structure of the workflows. And, while individual communities will have their own sets of workflows and objectives, an over-arching plan is needed to coordinate GEOSHARE’s strategy. One approach would be to focus on only a few high profile issues and the associated data bases. Another would be to bring in as much data as possible and hope that it will prove useful in informing decision making – even if the quality of some of these data is poor. Regarding quality, it is important that the domain experts put their assessments in the metadata information. The datasets and workflows should also be forward-looking – enabling the construction of ‘next generation’ models instead of simply improving on the last generation. A case in point was the need for intra-annual data on irrigation and cropping practices to improve future crop models.

An important theme cutting across this and other sessions was the immense value of tools posted on HUBZero in providing documentation of all steps in the development of a particular workflow, along with appropriate metadata. When combined with tools for viewing and comparing data, reporting key summary statistics, extraction of data and allowing for consistency checks and scrutiny of outliers (e.g., using box and whisker plots), this could be an important vehicle for data evaluation and endorsement.

Q3: Mechanics of endorsement: Who does the endorsement? How frequently, and in what form?

Participants see the Science Committee playing the lead role in determining data quality, based on the quality of the input data, the methods used to ‘fuse’ the various data sources and the meta-data provided. However, endorsement would be by the stakeholders represented on the Advisory Board.

Q4: Is broad stakeholder input needed? If so, how will this be collected and integrated as part of the endorsement process?

This issue was not addressed in detail, but it appears the stakeholder community could be quite broad. Focusing on a few high priority questions of interest to groups with funds will be important. The idea of soliciting input via an online survey was also raised.

Data Endorsement Breakout: Rapporteur Marc Imhoff

Q1: What is the value of GEOSHARE-endorsed geospatial data products for agriculture, food security and the environment? Who will benefit from such endorsement? Are there downsides from such endorsement?

The value endorsement depends on the user community.

The Science Community is best primary focus as the policy sector will ostensibly seek guidance from that community.

For the science community to be interested, data content, version tracking, and well described quality indicators will be of highest importance.

Efforts like IFPRI's assessment and rectification of FAO data is good example.

Metadata will be very important for all data and may range from very quantitative assessments in some cases to qualitative general descriptions by identified experts on the process and their confidence level in the data.

- Q2: Different datasets and workflows may be suited for different purposes: Which of these is most pressing for GEOSHARE to achieve, and how might this be accomplished?

Identification of key questions linked with primary scientific and other stakeholder communities will be critical for establishing the structure of the workflows

There was much discussion around the different workflows. Each partner or node will have its own set of workflows to meet their objectives but an overarching plan is needed to coordinate the strategy of GOESHARE with respect to Endorsement especially with respect to content and quality. It was generally felt that the science committee with international representation would be best suited to establish protocols for handling the Q/A and metadata required to cover the different data types.

Two strategies were postulated (not necessarily mutually exclusive). One was to focus on only a few high quality high profile issues and data, another was to bring in in as much data as possible even if the confidence level is low as long as it is described (IFPRI's North Korea project was an example).

Geoshare nodes can "evaluate" their own data and models based upon an agreed set of protocols for filling out metadata information so that large volumes of data can be included with transparency in their evolution and expert-based confidence assessments. It is important that the domain experts put their assessments in the metadata information.

The question arose as to whether or not a diagnostic process for evaluating data could be implemented.

There was also concern about missing data. Key data are still missing e.g., soils maps that contain sufficient information to assess soil performance and climate impacts and resiliency and ground water data. Identifying key missing data will be valuable as a first step followed by gap filling.

The workflows should also be forward looking and set the stage for the next generation of data, models and tools. For example, data that capture temporal changes or time series like intra annual crop maps and yield estimates will be important.

Establishing standards to follow which are evaluated periodically will be important.

- Q3: Mechanics of endorsement: Who does the endorsement? How frequently, and in what form?

Identification of key questions linked with primary scientific and other stakeholder communities will be critical. GEOSHARE will agree to provide data meeting high standards of Q/A and metadata transparency. Stakeholders can “endorse” the data.

This would need to be addressed by the science committee.

- Q4: Is broad stakeholder input needed? If so, how will this collected and integrated as part of the endorsement process?

This issue was not addressed in detail, but if the science community is the stakeholder it is pretty broad. Focusing on a few high priority questions of interest to groups with funds will be important.

Data Endorsement Breakout: Rapporteur Hermann Lotze-Campen

Q1: What is the value of GEOSHARE-endorsed geospatial data products for agriculture, food security and the environment? Who will benefit from such endorsement? Are there downsides from such endorsement?

- Not clear what endorsement exactly means; needs to be better defined; "validation" should be omitted
- Online discussion/review processes could be organized
- Develop "Rating system"; Platinum/Gold/... datasets in AgMIP (criterion: quality of indicators and purpose); Enable online discussions
- Minimum requirement: peer-reviewed publication with specific dataset
- Not too much responsibility should be put on researchers? ("data are always false"); "recommendations" should be sufficient; built-in diagnostics are needed; which horse for which course?
- Huge interest from private sector for endorsement (for regulatory requirements)
- "Official data" cannot be provided or endorsed by scientists

- BUT: Geoshare work should be targeted towards agreed-upon workflows (incl. confidence levels)
- Example: GTAP may be criticised, but at least it is internally consistent

Q2: Different datasets and workflows may be suited for different purposes: Which of these is most pressing for GEOSHARE to achieve, and how might this be accomplished?

- Land cover/use: as detailed as possible (available for countries like US, CAN,)
- Time series of products like Monfreda, Portman, Siebert, ... are needed; as timely as possible
- Long list of requests (see Niven's ppt)
- Much more resources are needed for setting up the necessary workflows!
- Information on management (and other socio-economic inputs) are needed
- Open Ag. data alliance – not present here, but should be involved
- Private companies should be able to provide selected datasets (new ways of cooperation needed)
- While huge amounts of distributed data are there – we still don't know the basics
- We are missing basic infrastructure for generating the necessary datasets
- INSPIRE (Europe) as an example
- "Distorted market": demand is there, but without an endorsed procedure/place

Q3: Mechanics of endorsement: Who does the endorsement? How frequently, and in what form?

No comments here

Q4: Is broad stakeholder input needed? If so, how will this be collected and integrated as part of the endorsement process?

- Open discussion (within boundaries) needs to be well organized in a simplified way (click buttons, ...); useful user interfaces
-

Additional comments by Niven Winchester on which data would be most valuable:

Thank you for inviting me to the GEOSHARE workshop last week. Now that I have had time to reflect on the meetings, I would like to note the items on my wish list that, in my opinion, would be of greatest benefit to the GEOSHARE community and would maximize exposure of the project.

- 1) Land area by vegetation time. In the breakout group that I was in, there was some disagreement about what level of disaggregation could be easily resolved. I think even a simple crop-pasture-

forest area disaggregation would be useful, with further disaggregation beyond these category a bonus (at least initially).

- 2) 2nd generation biofuel crop yields. In addition to facilitating the representation of these fuels in economic models, this would put GEOSHARE on the radar of oil companies and offer another (potential) funding source. I have been in touch with a group at the University of Illinois that is estimating global spatial yields for energy grasses and woody crops. Once their study is published, the authors have agreed to share their data with the Joint Program (which we will use to calibrate biofuels in the EPPA model), so I assume they would also be willing to upload their data to GEOSHARE.