Building bridges between domain scientists and machine learning experts:

The essential role of weather/climate scientists in machine learning collaborations

- **Imme Ebert-Uphoff** (Electrical & Computer Eng., Colorado State University)

Summary of experience based on collaboration with many great folks:

- **Yi Deng** (Climate Science, Georgia Tech)
- **Elizabeth Barnes** (Climate Science, Colorado State University)
- **Savini Samarasinghe** (Ph.D. student, Colorado State University)
- **Suzanne Pierce** (Hydrology, UT Austin)
- **Deanna Pennington** (Geology, UT El Paso)
- **Vipin Kumar** (Computer Science, U Minnesota)
- **Anuj Karpatne** (Computer Science, Virginia Tech)
- Others…

1st Workshop on Leveraging AI in the Exploitation of Satellite Earth Observations & Numerical Weather Prediction - Apr 25, 2019

Where I'm coming from

My research:

- **Causal discovery in climate science** (9 years): identifying potential cause-effect relationships from data.
- **ML for climate science** (5 years).

My community work – building bridges between ML folks and earth scientists:

1. **Climate Informatics community** (climateinformatics.org) - co-organizing annual workshop (NCAR).
2. **IS-GEO community** (IS-GEO.org).
   Intelligent systems for the geosciences (NSF-funded RCN)
3. Co-organizing sessions at AGU Fall meeting (ESSI, about 3 per year).

Note: I'm fairly new to weather / NOAA / AMS.

Combining ML/AI and weather/climate

ML / AI is emerging in climate & weather:

→ you saw all the reasons over the past 3 days.

If you are an expert in *both* **atmospheric sciences** *and* **machine learning** – that's great. Congratulations!

BUT: many of us only have expertise in *one* of these two areas

… then what?

Combining ML/AI and weather/climate

ML / AI is emerging in climate & weather:

→ you saw all the reasons over the past 3 days.

If you are an expert in *both* **atmospheric sciences** *and* **machine learning** – that's great. Congratulations!

BUT: many of us only have expertise in *one* of these two areas … then what?

Online learning resources are great…

… but finding a good collaborator is even better!

Topic requires *deep* collaboration – Why?

Earth science applications differ from typical ML applications.

Challenges include:

- Spatio-temporal structure;
- Heterogeneity in space and time;
- Multi-source, multi-resolution data;
- **Small sample size, lack of labeled data**.

  → Standard ML methods often not directly applicable.

  → Need to work closely together to adjust methods.

Advantage:

- **We have 100s of years of knowledge about underlying physical processes**!

  → Use that to compensate for lack of labeled data.

  → One more reason why weather/climate scientist is so important in collaboration!

Meet Peter and Andrea – Two companions throughout this discussion

Peter

Andrea

Cartoon guide: Ebert-Uphoff and Deng,
Three Steps to Successful Collaboration with Data Scientists,
AGU - **EOS** magazine, Aug 2017.

Meet Peter and Andrea – Two companions throughout this discussion

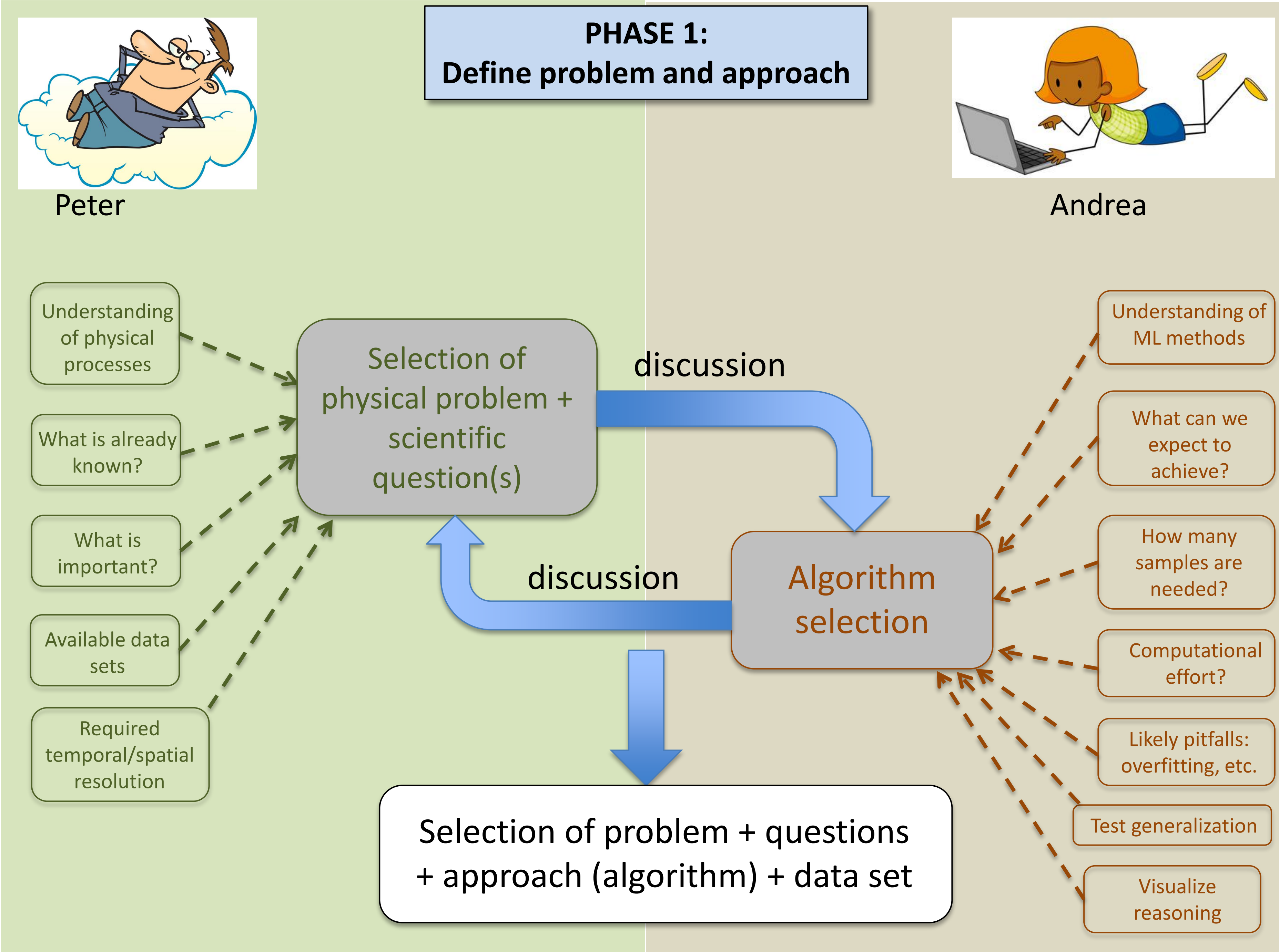
Peter

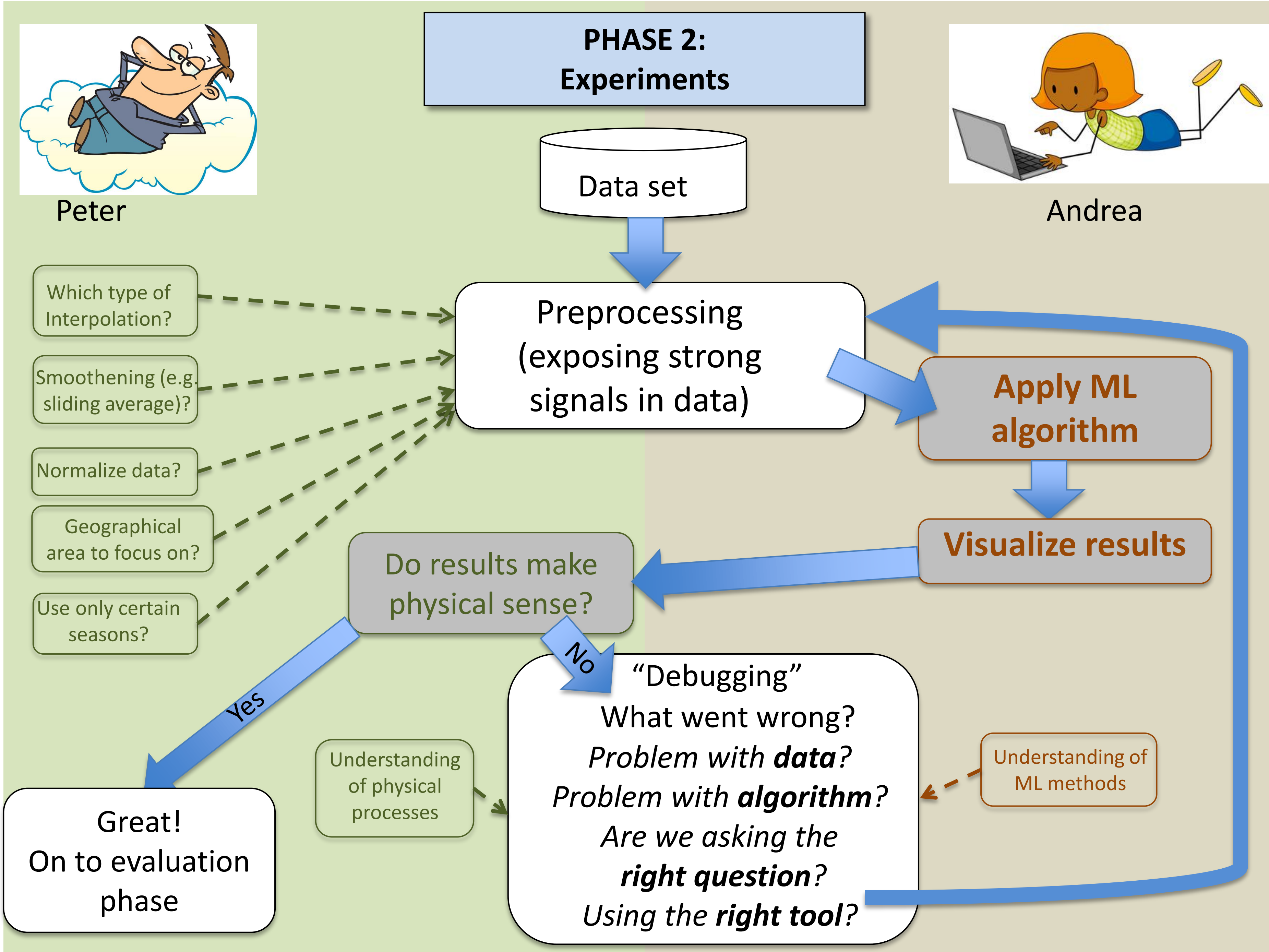
Weather / Climate scientist

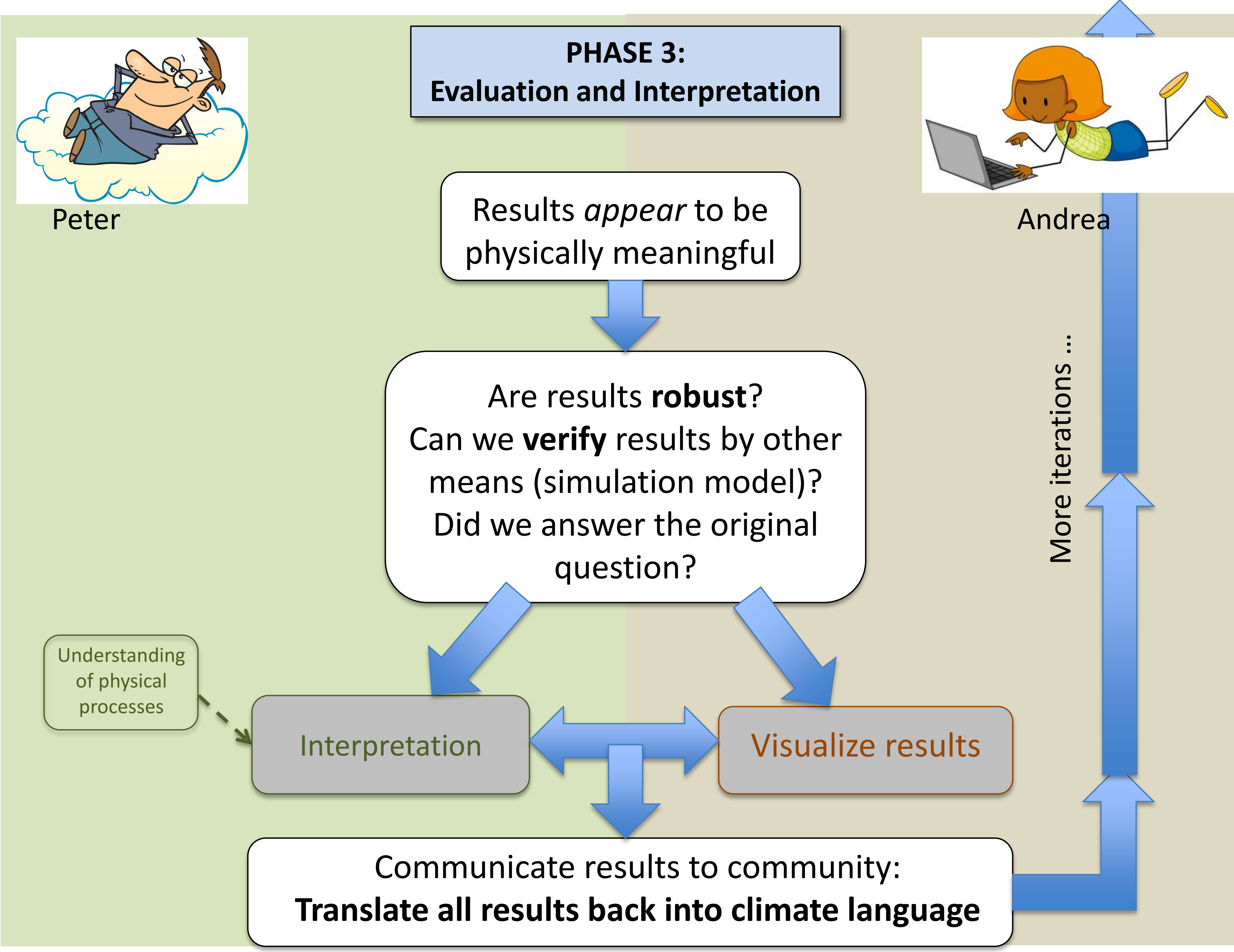
Andrea

Data scientist

Peter and Andrea want to learn to work more closely together.


PHASE 1:
Define problem and approach
Peter
Andrea
Understanding of physical processes
What is already known?
What is important?
Available data sets
Required temporal/spatial resolution
Selection of physical problem + scientific question(s)
discussion
discussion
Algorithm selection
Understanding of ML methods
What can we expect to achieve?
How many samples are needed?
Computational effort?
Likely pitfalls: overfitting, etc.
Test generalization
Visualize reasoning
Selection of problem + questions + approach (algorithm) + data set


PHASE 2:
Experiments
Peter
Andrea
Data set
Preprocessing
(exposing strong
signals in data)
Which type of
Interpolation?
Smoothening (e.g.
sliding average)?
Normalize data?
Geographical
area to focus on?
Use only certain
seasons?
Apply ML
algorithm
Visualize results
Do results make
physical sense?
Yes
No
Great!
On to evaluation
phase
Understanding
of physical
processes
"Debugging"
What went wrong?
Problem with data?
Problem with algorithm?
Are we asking the
right question?
Using the right tool?
Understanding of
ML methods


Peter
PHASE 3:
Evaluation and Interpretation
Andrea
Results *appear* to be physically meaningful
Are results **robust**?
Can we **verify** results by other means (simulation model)?
Did we answer the original question?
More iterations ...
Understanding of physical processes
Interpretation
Visualize results
Communicate results to community:
Translate all results back into climate language

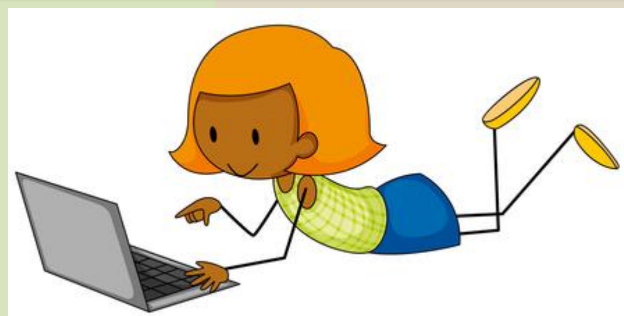
Observations:

1) Many tasks cannot be split into two separate parts that each person works on independently.
2) Many decisions must be made *together,* requiring both of their special knowledge.

Therefore:

1) Peter and Andrea cannot stay completely on their own side.
2) Each person needs to have a basic understanding of the thinking process of the other person.
3) Each person must be willing to teach / learn some basic vocabulary and tools.
4) Constant feedback from both sides is essential. **Talk to each other, talk, talk, then talk some more!**

Climate/Weather science



Data science

Learning to work together – in spite of vocabulary, culture, etc

Peter and Andrea have a topic for collaboration in mind.

Technique 1: The Interview

Step 1: Andrea interviews Peter:

> "Peter – Tell me all the basics about the physical problem. But in plain English, please!"

- Andrea takes notes. Types them up. Sends them to Peter for checking.
- Peter sends back corrections – until they both agree.

→ Yields problem formulation they both agree on.

→ **Document is in a language they both understand.**

→ Co-created first boundary object (crossing interdisciplinary boundaries).

If Peter just types up the problem → No check for understanding built in. Less learning taking place. No knowledge integration. No crossing boundaries.

Peter and Andrea want to learn to work more closely together.

Step 2: Reverse. Peter interviews Andrea about potential methods…

This interview phase nicely sets the stage for communication habits for entire collaboration.

It tests whether:

- **Both parties want to understand and learn about the other discipline;**
- **Both parties are willing to spend time explaining and listening;**
- **The communication works between the two - both can explain their own topics in plain English.**
- **CAUTION: If you can't make this phase work – you may want to look for a new collaborator!**
- *How strictly you follow this process depends on how far apart the parties are, and how well they communicate naturally.*

Learning to work together

Technique 2: Concept maps

- Helpful for larger teams.
- No time to discuss.
- Some References:

Pennington, D., Bammer, G., Danielson, A., Gosselin, D., Gouvea, J., Habron, G., Hawthorne, D., Parnell, R., Thompson, K., Vincent, S. and Wei, C., 2016.
The EMBeRS project: employing model-based reasoning in socio-environmental synthesis. Journal of Environmental Studies and Sciences, 6(2), pp.278-286.

M.L. Deaton, C.A. Wei, and Y.-C. Weng,
Concept Mapping: A Technique for Teaching about Systems and Complex Problems
https://www.sesync.org/concept-mapping-a-technique-for-teaching-about-systems-and-complex-problems

Foster these skills in the team

Source: Newell and Luckie, *Pedagogy for Interdisciplinary Habits of the Mind*, Conference on Interdisciplinary Teaching and Learning, 2012.

Interdisciplinary Habits of the Mind

Subset:

- **Set aside personal convictions;**
- **Strive for a feel of each discipline's perspective;**
- Embrace contradictions (ask how it can be both);
- **Strive for balance (among disciplinary perspectives);**
- **Don't fall in love with a solution until you understand the full complexity of the problem;**
- ***Value intellectual flexibility and playfulness.***

Helpful **Personal Qualities and Skills**

Foster these skills in yourself & Look for these skills in collaborators.

- **Communication skills, organizational skills;**
- **Broad interest**, **flexibility**, **creativity**, **openness**;
- Tolerance for ambiguity;
- **Transcendence of disciplines;**
- **Respect toward people, perspectives, and cultures;**
- Scientific skills for gathering, translating, analyzing, structuring, weighting and valuing, and synthesizing knowledge and information.

Source: Flinterman et al., *Transdisciplinarity: The New Challenge for Biomedical Research*, Bulletin of Science, Technology & Society, Vol. 21, No. 4, 2001.

Where to go from here?

Some suggestions (my 10c worth):

- **This workshop is amazing! Excellent job, excellent discussions. Again next year?**
- **Create interest group(s) / task force for selected topics**
- **Sample activities:**

1. Develop document summarizing ***Best practices for use of AI in weather/climate.***

   (Amy McGovern and I started planning that last night)

2. Develop **documented case studies –** showing what went right / wrong. (I want to interview Jebb & Christina for that.)

3. Work with AMS AI conference. Connect to CI, IS-GEO.

4. Provide contact point / place where people can connect / place to coordinate group activities. (Slack channel? Open Google docs on different topics?)

Research Suggestions

- **Improve transparency of ML algorithms**
  - New visualization techniques exist – use them! (Amy, Ryan)
  - Working group topic?

- **Merge physics into ML algorithms:**
  - At "feature selection" stage,
  - As constraints in optimization problems,
  - New field: Physics-guided machine learning (PGML) (Vipin Kumar, Anuj Karpatne).
  - Working group topic?