Biomedicine Breakout 2

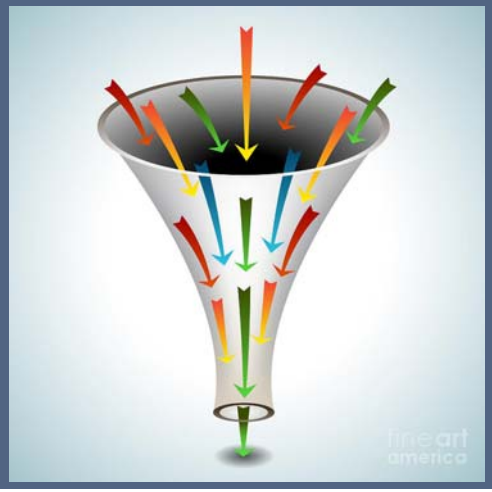
Stan Ahalt (UNC-CH, Data Translator)
Rafael Gonçalves (Stanford, CEDAR, Protegé)
Susan Gregurick (NIH, NITRD)
Charlotte Nelson (UCSF, SPOKE)
Dexter Pratt (UCSD, Cytoscape, NDEx)
Michael Witbrock (IBM Research, AI)

Ontologies

<u>Challenge for OKN</u>: If we want to develop a computable, intelligent knowledge representation network, we need to have a definable reference set….but the enemy of the good is the best.

<u>Potential Solutions</u>, short term 1-2 years

- **Corpus of Ontologies**, either all together in one place (equlv. to UICU for fist 100s webpages) OR smart APIs and stable location references.
  - **Computable NOT Clickable**
- **Humans should not (have to) create ontologies**



- Converging ontologies down to a minimally convergent set
- Extracting best, consistent information
- Create Computable representations for effective spiders

➢ **Even 50% of completeness is fine,** enemy of the good is the best, let's make incremental improvement.

➢ Minimal FTE level is needed, low hanging fruit.

Licenses, Provenance and Security

Challenge for OKN: Data, Information, Publications, Images….all of these have a variety of licenses, metadata, provenance.

Potential Solution: Sorter term 2 years
Maintain and relational link to knowledge representations (eg data) for license information, metadata, publications, provenance..

Grand Challenge for OKN.V2: Human-centered data is a treasure trove for meaningful inferences, how do we know where it is, what it is, format….and compute on it

Potential Solution, Longer term

- We need to have a way to compute on **data behind a firewall**. We need a schema, a mechanism, maybe infrastructure.
- This schema needs to be consistent with what is developed for open data/knowledge

- Push algorithms to the data behind a firewall, sandboxing

Building Buy-In

Challenge for OKN: As important to having the ontologies, data… in a converging, computing scheme, understanding the relationships and reliability of those entities is equally important.

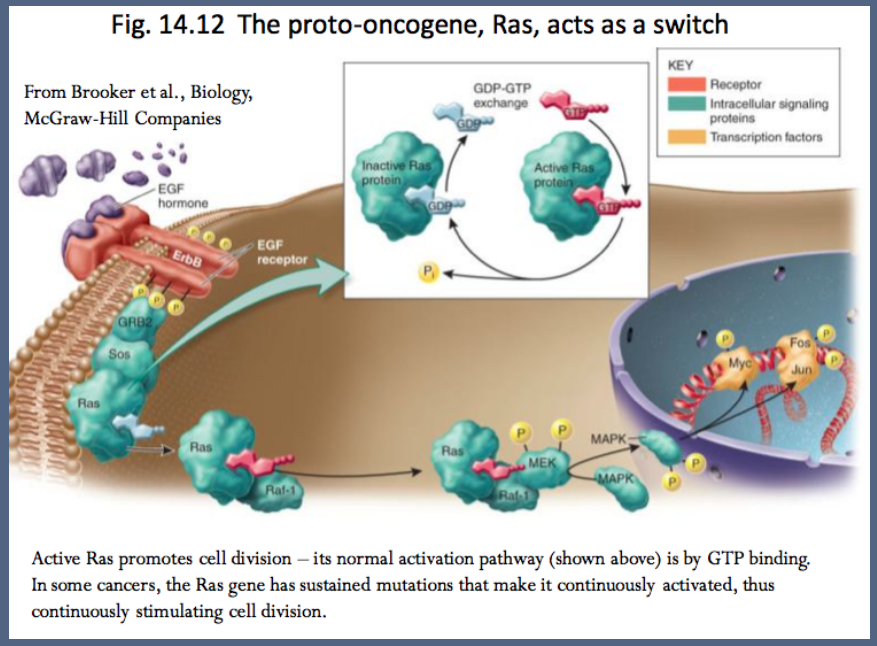
Potential Solution: Midterm 3-5 years

- Develop embedding relationships with probabilities within the knowledge representations
- Develop knowledge relationships that can be evaluated against experiments, and reconcile the knowledge representations (aka not static)
- Understanding Utility and Relative Utility, useful and believable for community buy-in

Use Cases

- Help biologists use a large(r) corpus of knowledge to query what types of experiments can be done and at what costs.

- Offer Clinicians alternative treatment pathways with some sense of reliability or potential outcomes and risks (based on knowledge inferences)



<u>Today:</u> RAS-Machine—Extract facts from literature and assemble onto signaling models for of RAS for RAS-related cancers.

- Updated daily
- Easy to build evaluation models and usability tests

John Bachman, Peter Sorger