



Grids and heterogeneity in bioinformatics

Martin Senger
martin.senger@gmail.com

What is this about?

- This is not a real presentation...
 - because you will not learn much
- This is rather a reminder
 - that the term GRID is still overloaded
 - and that various GRIDs are already used in bioinformatics without knowing (much) about it
- This may also leak how little I know about GRIDs
- ...and it is short (just seven slides)

GRID is about sharing resources

- the problem starts because we do not define the term “resources” uniformly
- the most visibly shared resources is the Internet itself – and still one would hesitate to call it GRID
 - why?
 - perhaps because such GRID is mostly accessible by human beings, not automated enough
 - btw, which may change with the advent of Semantic Web
- So?

GRID is about sharing resources and accessing them by programs

- here another problem starts
 - in order to access *things* by programs we need some standards (programs are stupid and there are too many of them)
 - the various resources may also need access to common datasets (BLAST databases?)
 - that may be far from easy to achieve efficiently enough
- So?

GRID is about sharing resources and accessing them by programs in a unified way

- well, how does it fit into bioinformatics domains?
- bioinformatics is both uniform and not uniform
 - it uses high-throughput clusters significantly uniform
 - and a plethora of scripts and do-it-fast programs
 - it uses and creates hundreds of databases, not necessarily in a unified way

Approaches in bioinformatics GRID

- Use as much as possible the “hidden” GRIDS that are already around us
 - clusters with LSF and other queuing systems
 - the “real” GRID environments (globus etc.)
- Add layers allowing to simulate the “unified way” (it seems like it is the same but it is not, but who cares?)
 - Examples...

"Applications GRIDs"

- Taverna
 - web services - but not only
 - a GRID with 1000+ nodes
- BioMoby
 - web services – but not quite
 - a GRID with 200+ nodes
- W3H
 - Web-based task manager
 - Author is here! (Are you, Peter?)
- Some of these "nodes" are the whole GRIDs on their own (using LSF etc.)



End of a reminder...

Thanks.