

Peer Review of Scientific Data

Great discussion tonight. I think the panel format is great for BESSIG.

I was thinking that, like scholarly publications, data files are bare, static entities. Both should be encapsulated in more intelligence. In the Object Oriented paradigm data is protected, encapsulated by software, with methods which provide access to the data inside the object capsule. All data could be Object Oriented, perhaps at the file level. All such OO files could be stored in OO filesystems and OO databases. There could be a base interface that is common to all such OO files. These common methods could be stuff like give me your DOI, give me your provenance. When the software is too tightly bound to the data, unlike the approach in which you distribute bare files and require software tools to inspect the data, you can do something like embedding a virus in the data, which is activated whenever the data object is accessed. In OO, when an object is instantiated, its constructor is called. This could be the virus, which does a number of things, like notify you that your data is being accessed. You would always know where your data goes and who uses it because it tells you. Otherwise, you rely on people to cite it or have to search publication databases to discover who mentions your data.