

DSpace * DS-638 check files on input for viruses, and verify file format

JIRA Reference: <https://jira.duraspace.org/browse/DS-638>

Proposed: "This patch uses JHOVE to provide rough-and-ready format checking by identifying that the file/bitstream extension matches formats verifiable by JHOVE. (Currently DSpace accepts a deposit's file extension as gospel, so a user could tack a ".txt" extension onto a GIF and DSpace would assign the incorrect format to the file based on that incorrect extension.) This patch also also contains code to check the file for the presence of viruses."

DCAT review: This patch seems to be doing two things: a) integrate with JHOVE, something that would be of strong interest for any repository that aim to preserve its contents in the long term and b) use virus checking tools (based on ClamAv) as part of that process. One can imagine that it would be interesting to have the virus tools without using the JHOVE package, so it may be worth exploring separating these? It may well be that it would be useful to encourage a community discussion about what tools would be useful, now that the curation framework is there (although I haven't had the time to check what it actually does).

The ticket is already assigned to Richard Rodgers as he will need to assess how this would work with the new curation framework that came with 1.7.

DCAT initial assessment: Relevant; Medium-High or High priority

Next steps: Initially it would be useful to check with Richard Rodgers what his take on this is (which I'm happy to do). Also, as Jim is also from Michigan, and the proposal seems to originate there, he may be able to provide more detail?

- If you agree with the above assessment and have no additional comments, you can simply respond with a +1.
- If you disagree but have no comments, a -1 works, and if you have no opinion at all, 0 is fine. (And encouraged, since that means we know you've had a chance to weigh in.)
- If you *do* have comments or other ideas, you're not limited to the numbers, of course. So please do share your thoughts!