



CSC

ICT Solutions for
Brilliant Minds



PIDit: kolme teesiä

Martin Matthiesen, CSC

PID-verkosto 10.9.2020



1 PID on lupaus.

- Älä luo PIDiä, jota et voi ylläpitää viiden vuoden päästä.
- Vähemmän on enemmän: 1000 hyvin ylläpidettyä PIDiä on parempi, kuin 25 tuhatta PIDiä, jotka ovat lähes mahdottomia ylläpitää.

2 Vain PIDit metadataan ovat välttämättömiä ja viittauskelpoisia.

- PID suoraan datalle ilman kontekstia ei sano mitään, ei datasta, ei lisensseistä.
- Metadatatista taas voi huoletta viitata URLilla dataan, koska PID mahdollistaa, että käyttäjät (ja koneet) tulevat PIDin kautta, ja jos metadata on standardisoitu, myös koneet pääsevät datalle automaattisesti käsiksi.

3 Varaudu muutoksiin ja hyväksy, että saatat rikkoa FAIR-periaatteet.



- Pysyvä datasetti ei ole niin pysyvää, eikä jokaista versiota kannattaa loputtomasti tarjota.
- Esimerkki: 250:n Gt zippi, jossa on zipin sisällä olevassa hakemistonimessä kirjoitusvirhe. Virheen voi korjata ja merkata metadataan, mitä on tehty, mutta FAIR vaatisi, että vanha zip pitäisi olla saatavilla ikuisesti.



Kiitos!



facebook.com/CSCfi



twitter.com/CSCfi



youtube.com/CSCfi



linkedin.com/company/csc--it-center-for-science



github.com/CSCfi