

Time Machine RFC ja GitHub – Määrittely-dokumenttien kirjoittaminen kollektiivisesti



Juha
Henriksson
Arkistonjohtaja
Musiikkiarkisto
RFC Editor
Time Machine

Time Machine -hankkeessa käytetään Internet-protokollien laatimisesta tuttua Request for Comments -prosessia, kun kirjoitetaan kollektiivisesti hankkeeseen liittyviä määräytyksiä. Time Machinen RFC-dokumenttien hallinta tapahtuu GitHub-palvelun kautta. Tämä artikkelin aiheena on Time Machinen RFC-prosessi ja GitHub, mutta toivottavasti se antaa samalla yleisempiäkin ideoita siihen, miten laajoissa hankkeissa dokumentteja voidaan tuottaa kollektiivisesti.

Time Machine

Time Machine on Euroopan ja maailmankin mittakaavassa ainutlaatuisen kunnianhimoisen ja laaja hanke, jonka tavoitteena on kehittää innovatiivisia ratkaisuja, joilla Euroopan arkistoissa, kirjastoissa ja museoissa säilytettävät kulttuuriperintöaineistot voidaan saattaa laajassa mitassa digitaaliseen muotoon. Digitoituja aineistoja voidaan tutkia ja visualisoida uusilla tavoilla esimerkiksi

tekstin- ja hahmontunnistusta sekä 3D-mallinnusta hyödyntäen.

Time Machine Organisaatioon¹ (TMO) on liittynyt yli 600 jäsentä, joiden joukossa on yliopistoja ja tutkimuslaitoksia sekä kulttuuriperintöorganisaatioita, mutta myös monia yrityksiä. Myös Suomessa on oltu aktiivisia, sillä maassamme on jo yli 40 TMO-jäsentä.

Time Machineen voi tutustua tarkemmin organisaation verkkosivuilla², mutta hanketta on käsitelty myös aikaisemmissa *Faili*-lehdissä (4/2019, 1/2020, 3/2020 ja 2/2021). Lisäksi hankkeesta – erityisesti Suomen näkökulmasta – kertovat laati-
mani sähköiset kirjat *Time Machine Suomessa* ja *Time Machine in Finland*, jotka on julkaistu avoimesti Musiikkiarkiston Open Access -sivuilla³.

Request for Comments

Time Machine on erittäin laaja kokonaisuus, jossa on mukana paljon erilaisia toimijoita. Miten voidaan varmistaa riittävä yhteentoimivuus kaikkien hyvinkin moninaisten teknologisten vaihtoehtojen välillä, joita yli 600 Time Machine -organisaatiossa on kehitetty tai tullaan kehittämään?

Time Machinassa asia on ratkaistu siten, että määritysten avoimen kehittämisen ja arvioinnin tueksi on otettu käyttöön Request for Comments -prosessi, joka on alun perin kehitetty Internet-protokollien laatimiseen. Time Machinassa on käytössä hieman muunneltu RFC-prosessi, jota on muokattu hankkeen tarpeita varten.

Time Machinen Requests for Comments -dokumentit⁴ ovat vapaasti saatavilla olevia julkaisuja, joissa määritellään hankkeen perusarkkitehtuuri sekä eri komponentteihin liittyvä säännöt ja suositukset. RFC-prosessin periaatteisiin kuuluvat:

- *Esteettömyys.* RFC-dokumentit ovat vapaasti ja veloituksetta käytettävissä.
- *Avoimuus.* Kuka tahansa voi kirjoittaa RFC-dokumentin.
- *Tunnistettavuus.* Jokaisella RFC-dokumentilla on yksilöllinen tunnus ja versionumero.
- *Inkrementalismi.* Jokaisen RFC-dokumentin tulee toimia sekä itsenäisenä dokumenttina että rakennuspalikkana muille dokumenteille.
- *Standardointi.* RFC-dokumenttien tulisi pyrkiä käyttämään standardoituja termejä, jotta

¹ <https://www.timemachine.eu/time-machine-organisation/>

² <https://www.timemachine.eu>

³ <https://www.musiikkiarkisto.fi/oa/kirjat.php>

⁴ <https://www.timemachine.eu/requests-for-comments/>

varmistetaan määritysten selkeys ja yksikäsitteisyys.

- *Laajuus.* RFC-dokumentit on tarkoitettu käytännön ongelmien ratkaisemiseen. Ne eivät ole tieteellisiä julkaisuja eivätkä välttämättä sisällä kokeellisia todisteita. RFC:t kattavat teknisen infrastruktuurin lisäksi Time Machinen tietostandardit, oikeudelliset puitteet sekä arvot ja periaatteet.
- *Itsemäärittävyys.* RFC:t ovat tärkein prosessi Time Machinen infrastruktuurin ja toimintaprosessin luomiseen, mutta sen lisäksi RFC-dokumenteissa määritellään myös se prosessi, jolla RFC-dokumentteja luodaan, arvioidaan, hyväksytään ja julkaistaan.

Time Machine käyttää RFC-prosessissa GitHub-palvelua, sillä se tukee

dokumenttien kollektiivista kirjoittamista. Sekä RFC-dokumenttien luonnokset että varsinaiset julkaistut dokumentit ovat GitHubin RFC-varastossa⁵.

Git ja GitHub

Mitä ovat Git ja GitHub? Git⁶ on ilmainen avoimen lähdekoodin ohjelmisto, jonka avulla voidaan seurata muutoksia ja koordinoita työtä ohjelmoijien kesken, kun lähdekoodia kehitetään kollektiivisesti useiden ohjelmoijien toimesta. Gitin tavoitteena on nopeus ja tietojen eheys sekä tuki hajautetuille, epälineaarisille työnkuluille, joissa voi olla jopa tuhansia rinnakkaisia haaroja.

Gitin laati alun perin suomalainen Linus Torvalds vuonna 2005 Linux-ytimen kehittämiseen. Toisin kuin

⁵ <https://github.com/time-machine-project/requests-for-comments>

⁶ <https://git-scm.com>

useimmat asiakaspalvelinjärjestelmät, jokainen Git-hakemisto muodostaa kyseiseen tietokoneeseen täysimittaisen arkiston, joka sisältää täydellisen historian ja täydet versionseurausominaisuudet, jotka eivät riipu verkkoyhteydestä tai keskuspalvelimesta.

GitHub⁷ on vuonna 2008 avattu verkkopalvelu, joka tarjoaa käyttöliittymän ja tallennuskapasiteettia Gitillä hallittuihin tietovarantoihin. GitHubin ytimessä on ohjelmakoodin pilvitalennus, mutta sen ympärille on rakennettu useita työkaluja ja ominaisuuksia, jotka helpottavat ohjelmistokehitystyötä. Näitä ovat esimerkiksi:

- Virheiden seuranta, johon on tarjolla esimerkiksi labelit, virstanpylväät (milestones), vastuunjako ja hakukone.

⁷ <https://github.com>

time-machine-project / requests-for-comments Public

Code Issues Pull requests Actions Security Insights

master requests-for-comments / files / releases / RFC-0002 / RFC-0002.md Go to file

yngwi Remove circular dependency from RFC-0002 Latest commit 1243a89 on 7 Jun History

1 contributor

1701 lines (1067 sloc) 43.3 KB

title	subtitle	author
RFC on RFC Tree	Time Machine RFC-0002	François Ballaud Frédéric Kaplan Isabella di Lenardo Kevin Baumer Daniel Jeller

Motivation

The process for writing RFCs requires long-term planning instruments which will enable writers to not only access constantly updated versions of current RFCs but also to have a view on future planned RFCs. The RFC Tree as described in this RFC will provide an up-to-date description of all planned RFCs, including a short textual motivation for each and listing their dependencies with other RFCs.

Time Machinen RFC-dokumentteja hallitaan GitHubissa. Kuvakaappaus: Juha Henriksson.



Time Machinen RFC-prosessin teknisestä toteutuksesta vastaa Chief Technical Officer Daniel Jeller.

Kuva: Caroline Maximoff.

- Pull Requests -toiminto, jolla hallitaan projektin eri haarojen yhdistämistä. Työkalun avulla voidaan ehdottaa muutoksia, jolloin nimetyt henkilöt voivat nähdä pyydettyjen muutosten erot ja hyväksyä ne.
 - Projektiin liittyvää dokumentaatiota voidaan hallita GitHub Wikissä tai erilaisissa Markdown-formaatissa olevissa tiedostoissa (esimerkiksi README).
 - GitHub Pages, joka on GitHubin tarjoama hosting-palvelu projektin dokumentaatiolle, käyttäjien blogeille ja jopa kokonaisille kirjoille.
 - GitHub kokoaa myös erilaisia tilastoja muun muassa projektin osallistujista ja nopeudesta.
- GitHub on ilmainen julkisesti nähtävillä oleville tietovarastoille. Yksityiset tietovarastot olivat aikaisemmin maksullisia, mutta vuoden 2019 alusta lähtien kaikki ovat voineet käyttää yksityisiä tietovarastoja rajoitetuilla ominaisuuksilla.
- GitHubia käytetään hyvin yleisesti avoimen lähdekoodin ohjelmistokehitysprojekteihin. Kesällä 2022 GitHubia käytti yli 80 miljoonaa kehittäjää ja yli 200 miljoonaa tietovarastoa. GitHub on ollut Microsoftin tytäryhtiö vuodesta 2018 lähtien.

RFC Editor

Time Machinen RFC-prosessin hallintaan liittyvästä teknologiasta vastaa TMO:n Technical Officer Daniel Jeller. Hän on itävaltalainen historioitsija ja tietotekniikan asiantuntija, joka toimii myös digitoinnin ja tietotekniikan johtajana ICARUS:issa, kansainvälisessä arkistotutkimuskeskuksessa Wienissä.

Olen itse toiminut helmikuusta lähtien osa-aikaisesti Time Machinen RFC-dokumenttien päätoimittajana eli RFC Editorina⁸. Työskentelen yhteistyössä Time Machinen johtoryhmän ja toimihenkilöiden kanssa,

⁸ <https://www.timemachine.eu/team/juha-henriksson-2/>

ja apunani on myös RFC-toimituskunta.

RFC Editorin tehtäviin kuuluvat:

- RFC-dokumenttien kirjoittajien rekrytointi
- RFC-kirjoitusprosessin ohjaaminen
- RFC-toimituskomitean kokoaminen ja hallinnointi
- Avustaminen RFC-aiheisten työpajojen järjestämisessä ja toteuttamisessa

RFC-julkaisuprosessi

Time Machinen RFC-dokumentit eivät ole tieteellisiä julkaisuja, vaan ne on tarkoitettu käytännön ongelmien ratkaisuun. RFC-julkaisuprosessi kuitenkin muistuttaa tieteellistä julkaisemista, sillä se perustuu vertaisarviointiin.

Myös itse RFC-julkaisuprosessi määritellään kollektiivisesti, sillä sekin on kuvattu RFC-dokumenteissa. Kaikkien RFC-dokumenttien muodostama kokonaisrakenne on määritelty niin sanotussa RFC-puussa (RFC Tree, RFC-0002), julkaisuprosessi on määritelty dokumentissa RFC-0001, ja myös päätoimittaja ja toimituskomitea määritellään omassa RFC-dokumentissaan (RFC-0004).

RFC-päätoimittaja ja/tai toimituskomitea nimittää kirjoittajat jokaiselle RFC-puussa määritellylle dokumentille. Vaihtoehtoisesti asiasta kiinnostuneet voivat ottaa myös itse ottaa yhteyttä päätoimittajaan ja tarjoutua kirjoittamaan tietyn

RFC-dokumentin – siinäkin tapauksessa, ettei sitä ole määritelty RFC-puussa. Mikäli haluaa päästä vaikuttamaan Time Machinen sisältöön ja osallistua RFC-dokumenttien kirjoitusprosessiin, kannattaakin olla yhteydessä allekirjoittaneeseen. Suomalaiset asiantuntijat ovatkin jo osallistuneet muutaman uusimman RFC-dokumentin kirjoittamiseen.

Kun kirjoittajat ovat saaneet valmiiksi luonnoksen, päätoimittaja ja/tai toimituskunta tarkistaa sen, jonka jälkeen luonnos lähetetään tarvittaessa muille asiantuntijoille arvioitavaksi. Mikäli luonnoksessa on puutteita, se lähetetään takaisin kirjoittajalle korjattavaksi. Hyväksytyt RFC-dokumentit julkaistaan GitHubissa, ja niistä kootaan myös vuosittain PDF-muotoinen RFC-kirja⁹.

Kokemuksia

Olen itse ehtinyt kirjoittaa tähän mennessä muutaman RFC-dokumentin. Oman kokemukseni mukaan muutosten ja versioiden hallinta sekä uusien, yhtenäisten kokonaisuuksien lisääminen sujuu GitHubin kautta selvästi helpommin kuin että kollektiiviseen kirjoittamiseen käytettäisiin vaikkapa Google Drivea.

GitHubissa työn alla olevasta dokumentista luodaan oma haara, johon voidaan sitten tehdä muutoksia ja lisäyksiä ilman että itsellään työstettävänä oleva dokumentti muuttuu sinä aikana toisten kirjoittajien toimesta. Kun omat lisäykset ja muutokset ovat valmiit, tehdään

pull request, jolloin järjestelmä tutkii kaikki mahdolliset ristiriidat eli onko joku toinen muuttanut samaan aikaan samoja asioita jotenkin toisin, jonka jälkeen ristiriidat voidaan ratkaista halutulla tavalla.

GitHubin kautta vältetään esimerkiksi se, että kaksi kirjoittajaa työstää samaa kohtaa jatkuvasti ”edestakaisin”. Muutosten ja versioiden hallinta on muutenkin hallittua ja hyvin dokumentoitua.

Toisaalta GitHubin käyttö on jonkin verran teknisesti haastavampaa kuin esimerkiksi Google Docsin, mutta ainakin itse pääsin sen toiminnoista varsin nopeasti perille. Aionkin harkita vakavasti GitHubin käyttöä Time Machinen lisäksi muissakin hankkeissa, joissa on tarkoitus kirjoittaa kollektiivisesti vähänkin pitempiä ja sisällöltään haastavampia dokumentteja.

⁹ <https://github.com/time-machine-project/requests-for-comments/releases/tag/2021.01>