



DICE

Data Infrastructure Capacity for EOSC

SURF Data Repository

Long-term preservation of large-scale
research data sets



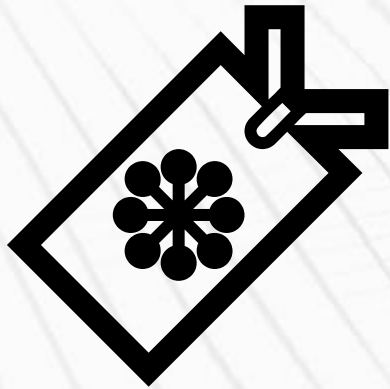
The DICE project has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 under Grant Agreement no. 101017207

Data Repository in a nutshell

- * B2SHARE-like repository service by SURF, the Netherlands
- * Optimized for large-scale research data sets
- * Domain-agnostic, supports any community

- * And:
 - * Self-service platform with roles
 - * Basket functionality for automated large-scale downloads
 - * Proximity with compute and other storage infrastructures
 - * Indexation by metadata catalogues, PIDs and DOIs
 - * REST API for automation

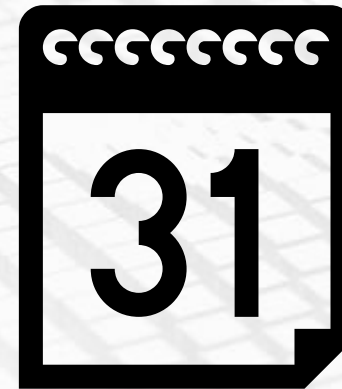
Data Repository in a nutshell



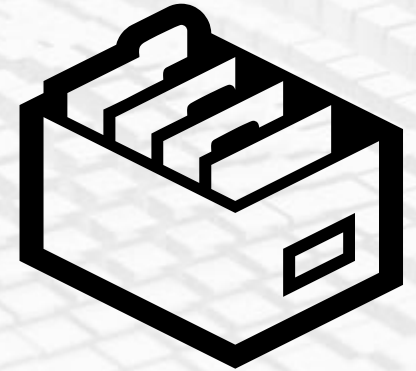
“ I want to **publish** my very large dataset in a **trustful** repository ”



“ I need **automated publishing** of my data but only to **specific** people ”

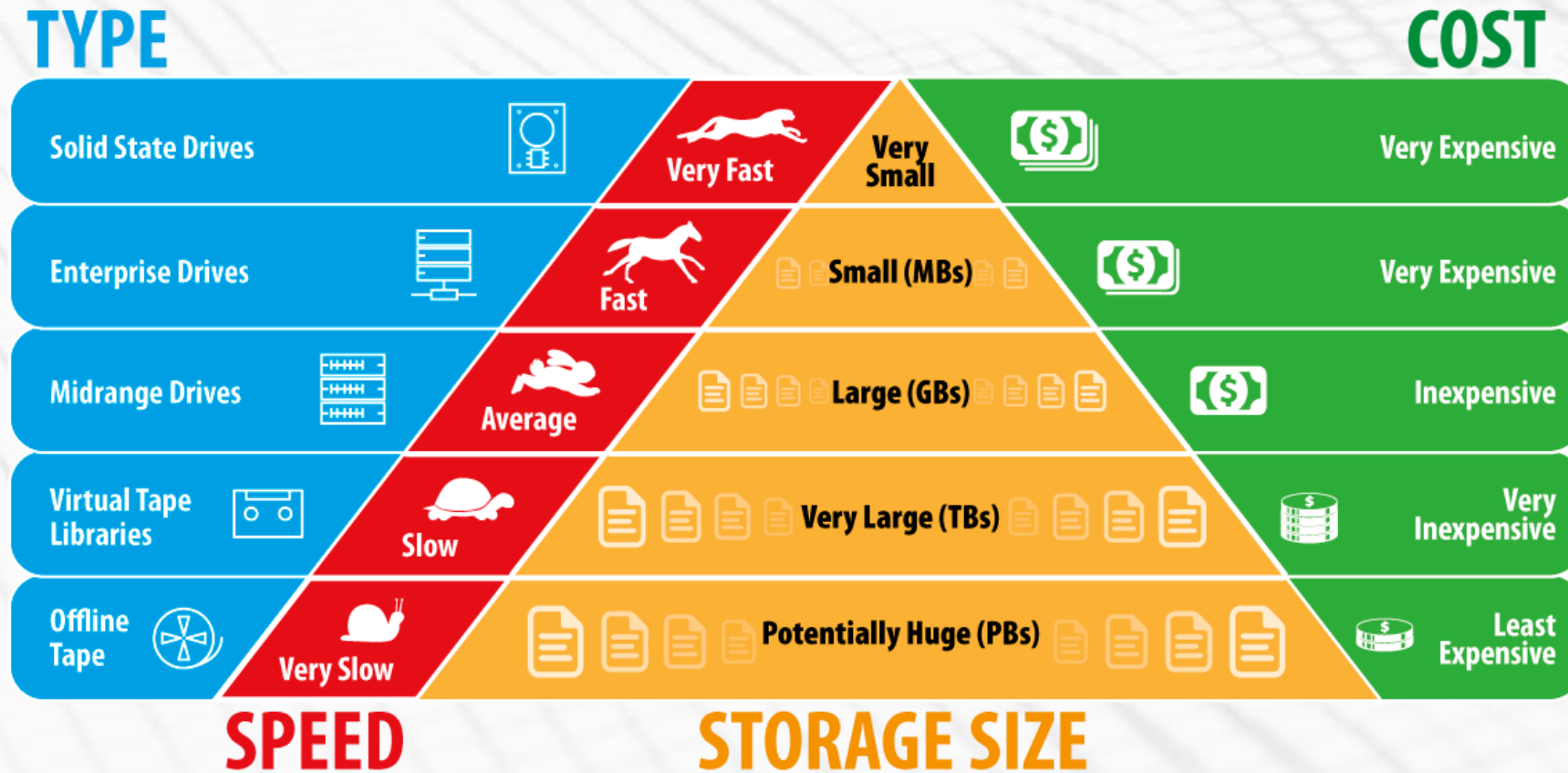


“ My dataset should have a **PID** and **metadata** attached, and should be publicly **available** for 10 years ”



“ I want to **manage** my own collections, schemas and publications ”

Storage trade-offs



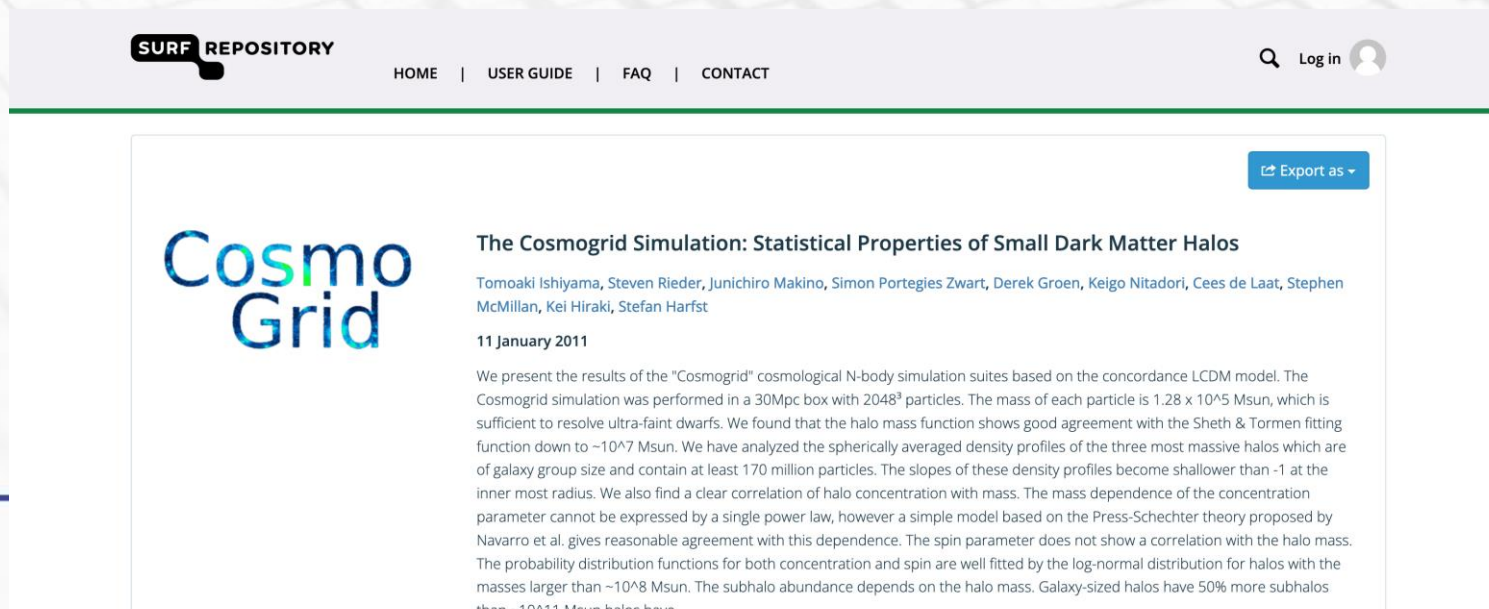
Data set access

- * Storage on tape-based infrastructure:
 - * Online and offline files => optimized for cost-efficiency
 - * Data not continuously directly available
- * Storage on fast horizontal scaling infrastructure
 - * Data directly available anytime
- * Metadata is always accessible

What is provided?

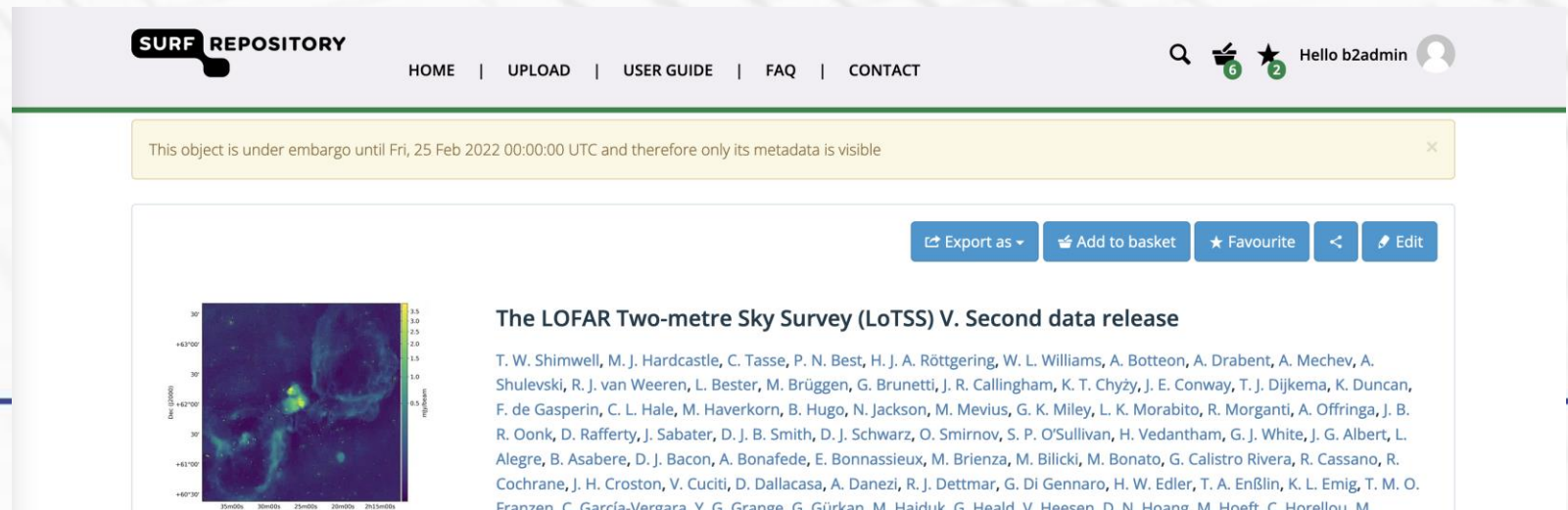
- * Structured approach: communities with schemas, collections and deposits
- * Community with policies and administrative roles
- * Landing page for every object
- * EPIC PIDs for objects and files
- * DOIs for collections and deposits (if required)

- * Publication of CosmoGrid cosmological simulations (2018)
 - * Data volume of 100 TB, already residing in Data Archive of SURF
 - * Structured using 1 collection containing >700 deposits
 - * Development of community & metadata schema
 - * Close collaboration with researchers



The screenshot shows the SURF Repository website. The header includes the SURF REPOSITORY logo, navigation links (HOME, USER GUIDE, FAQ, CONTACT), a search icon, and a Log in button. The main content area features the CosmoGrid logo on the left and a detailed article on the right. The article title is "The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos" by Tomoaki Ishiyama, Steven Rieder, Junichiro Makino, Simon Portegies Zwart, Derek Groen, Keigo Nitadori, Cees de Laat, Stephen McMillan, Kei Hiraki, and Stefan Harfst, dated 11 January 2011. The article text describes the results of the "Cosmogrid" cosmological N-body simulation suites based on the concordance LCDM model. It mentions that the simulation was performed in a 30Mpc box with 2048³ particles, with a particle mass of 1.28 x 10⁵ Msun. The article discusses the halo mass function, density profiles, and the correlation of halo concentration with mass. It also mentions that the mass dependence of the concentration parameter cannot be expressed by a single power law, but a simple model based on the Press-Schechter theory proposed by Navarro et al. gives reasonable agreement with this dependence. The spin parameter does not show a correlation with the halo mass. The probability distribution functions for both concentration and spin are well fitted by the log-normal distribution for halos with the masses larger than ~10⁸ Msun. The subhalo abundance depends on the halo mass. Galaxy-sized halos have 50% more subhalos than ~10¹¹ Msun halos have.

- * Publication of [LoTSS-DR2 research data set](#) (2022, *under embargo*)
 - * Data volume of 150 TB using two storage infrastructures
 - * Structured using 1 collection containing >800 deposits
 - * Development of community & metadata schema
 - * Close collaboration with researchers
 - * Development of new functionality as needed



SURF REPOSITORY HOME | UPLOAD | USER GUIDE | FAQ | CONTACT

Search 6 2 Hello b2admin

This object is under embargo until Fri, 25 Feb 2022 00:00:00 UTC and therefore only its metadata is visible

Export as Add to basket Favourite Edit

The LOFAR Two-metre Sky Survey (LoTSS) V. Second data release

T. W. Shimwell, M. J. Hardcastle, C. Tasse, P. N. Best, H. J. A. Röttgering, W. L. Williams, A. Botteon, A. Drabant, A. Mechev, A. Shulevski, R. J. van Weeren, L. Bester, M. Brüggen, G. Brunetti, J. R. Callingham, K. T. Chyży, J. E. Conway, T. J. Dijkema, K. Duncan, F. de Gasperin, C. L. Hale, M. Haverkorn, B. Hugo, N. Jackson, M. Mevius, G. K. Miley, L. K. Morabito, R. Morganti, A. Offringa, J. B. R. Oonk, D. Rafferty, J. Sabater, D. J. B. Smith, D. J. Schwarz, O. Smirnov, S. P. O'Sullivan, H. Vedantham, G. J. White, J. G. Albert, L. Alegre, B. Asabere, D. J. Bacon, A. Bonafede, E. Bonnassieux, M. Brienza, M. Bilicki, M. Bonato, G. Calistro Rivera, R. Cassano, R. Cochrane, J. H. Croston, V. Cuciti, D. Dallacasa, A. Danezi, R. J. Dettmar, G. Di Gennaro, H. W. Edler, T. A. Enßlin, K. L. Emig, T. M. O. Franzen, C. García-Vergara, Y. G. Grange, G. Gürkan, M. Haiduk, G. Heald, V. Heesen, D. N. Hoang, M. Hoeft, C. Horellou, M.

Gaining access

- * Virtual access via DICE until June 30th, 2023
- * Subsidized access for Dutch research institutes
- * Contractual access for any other research institute

Long-term preservation

- * Storage: bitwise preservation
- * Curation level: basic curation
 - * Secure storage of files and related data
 - * Pre-publication structuring, file format migration and schema building
 - * Annotation and basic checks on metadata
 - * Metadata versioning

Long-term preservation

- * Data Repository adheres to internal Long-term Preservation (LTP) Policy:
 - * Dual storage of all files
 - * Checksum and fixity checks
 - * Storage medium migration
 - * Continuous monitoring
 - * Data set content curation upon request



JOIN OUR COMMUNITY



DICEosc



[/company/diceosc](#)