



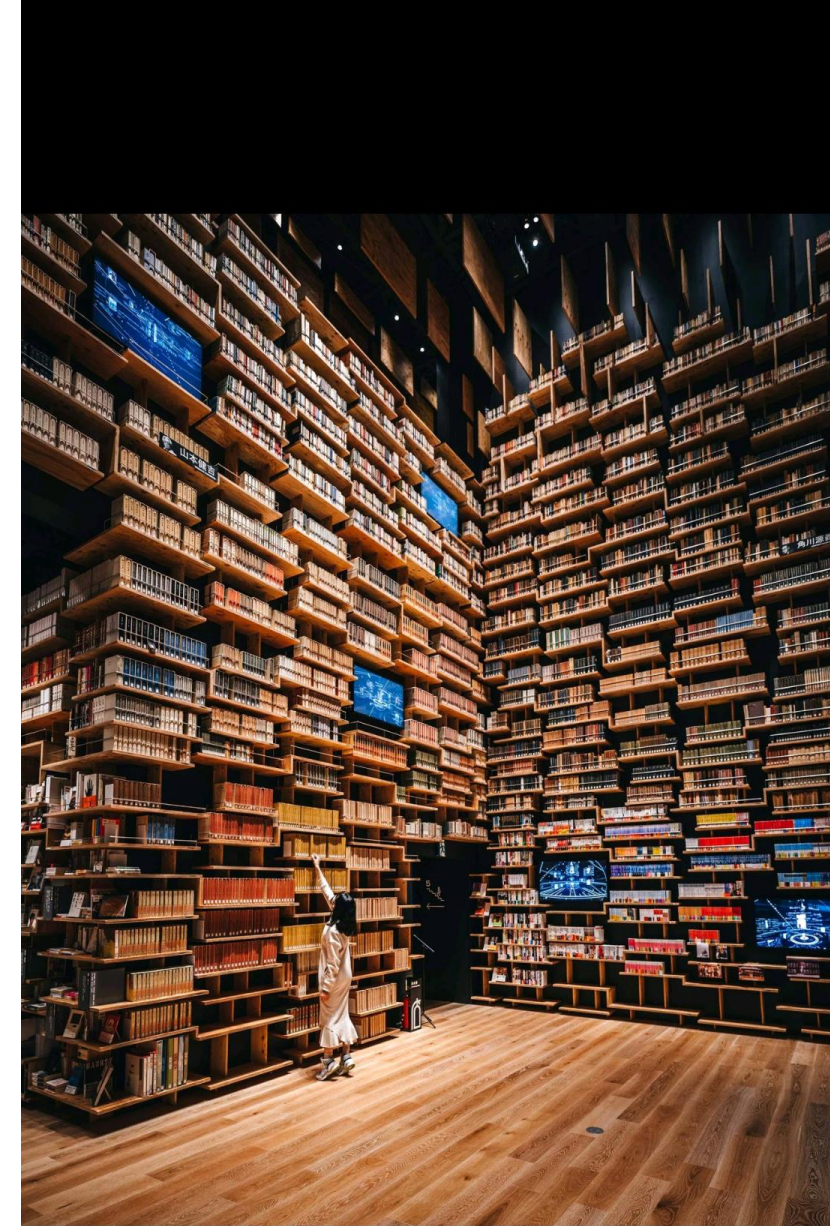
Dataset Exchange Working Group



Data Catalog Vocabulary -
DCAT

From 10,000 metres

- Catalogues are created to organize 'things' so that we know what is there and where to find it.
- A catalogue holds and organizes the basic information to find resources
- The W3C Data Catalog Vocabulary (DCAT) is a model for describing catalogues
- These catalogues are generally used for datasets, but there is scope for other resource types
- There are other vocabularies for describing catalogues of data - DCAT is meant to be an interoperability standard, a *lingua franca* that supports simple aggregation and federated query of catalogues.
- DCAT is frequently augmented by other vocabularies to provide an application profile



Origins of DCAT

- Designed at the Digital Enterprise Research Institute (DERI) in Galway, Ireland, and contributed to the W3C Government Linked Data Working Group (2011-2013)
- Developed during the first wave of government open data portals, with the purpose of providing a cross-platform standard, based on Semantic Web technologies
- Published in 2014 as a W3C Recommendation



DCAT 1 (2014)

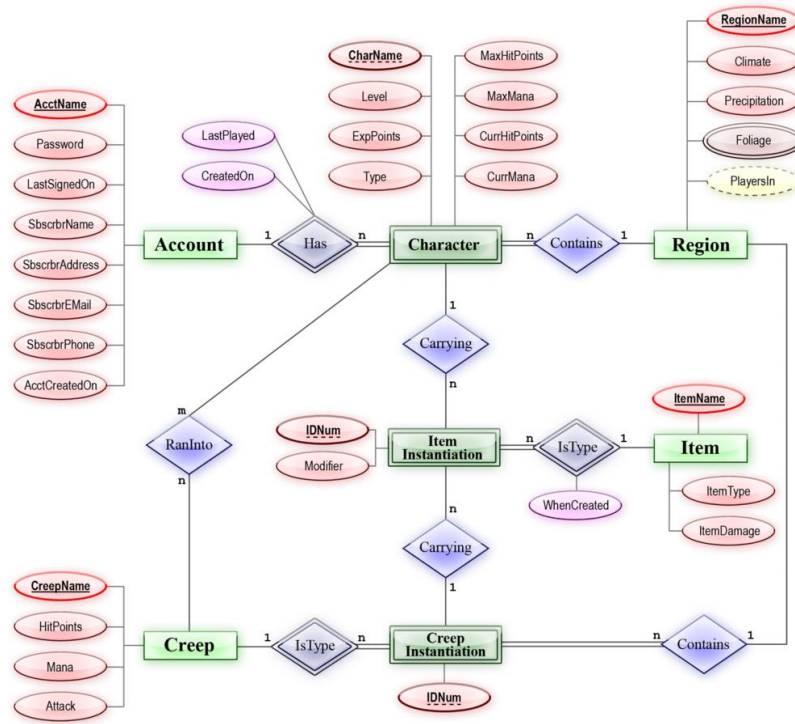
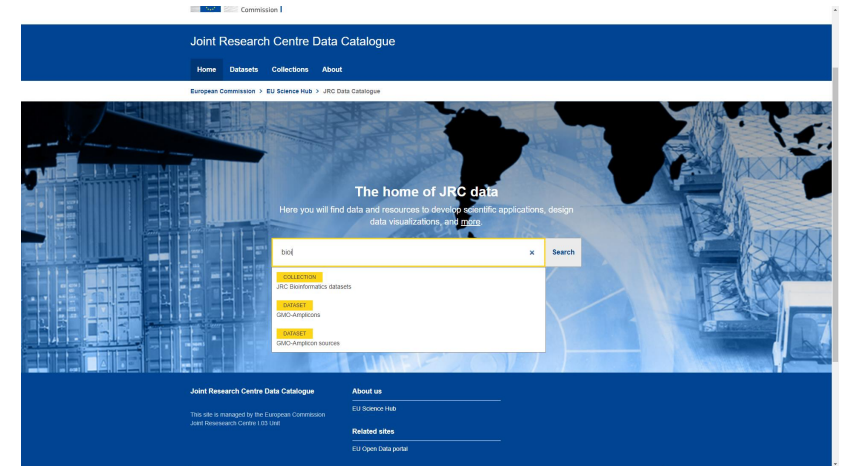
Catalog

The screenshot displays the Joint Research Centre Data Catalogue website. At the top, there is a blue header with the text "Joint Research Centre Data Catalogue" and a navigation menu with links for "Home", "Datasets", "Collections", and "About". Below the header, a breadcrumb trail reads "European Commission > EU Science Hub > JRC Data Catalogue". The main content area features a large background image of a warehouse with a satellite in the sky. A central graphic shows a speaker icon with sound waves and the text "The home of JRC data". Below this, a sub-header states: "Here you will find data and resources to develop scientific applications, design data visualizations, and more." A search bar contains the text "bio" and a "Search" button. The search results are displayed in a white box with a yellow border, listing three items: a "COLLECTION" for "JRC Bioinformatics datasets", and two "DATASET" entries for "GMO-Amplicons" and "GMO-Amplicon sources". The footer contains the site title "Joint Research Centre Data Catalogue", a management notice, and sections for "About us" (with links to "EU Science Hub") and "Related sites" (with a link to "EU Open Data portal").

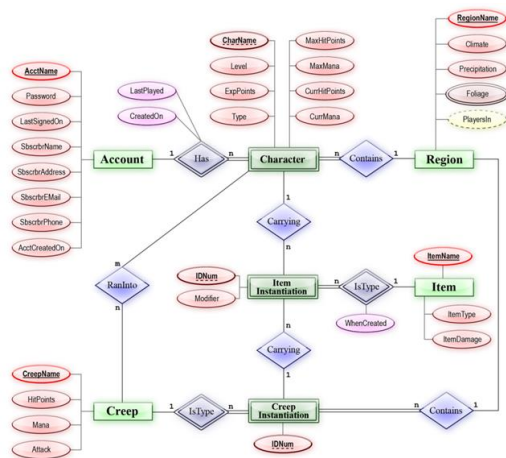
DCAT 1 (2014)

Catalog

Dataset



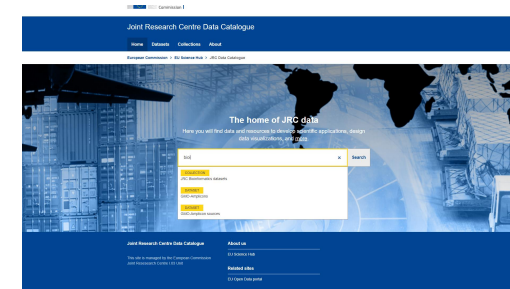
DCAT 1 (2014)



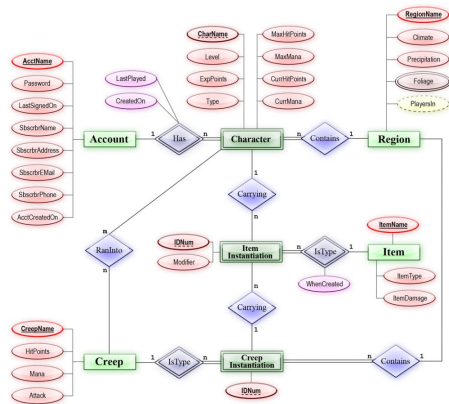
Catalog

Dataset

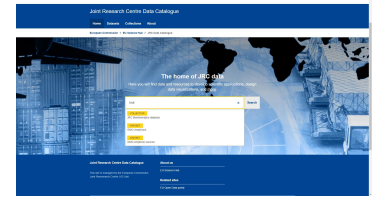
Distribution



DCAT 1 (2014)



Catalog



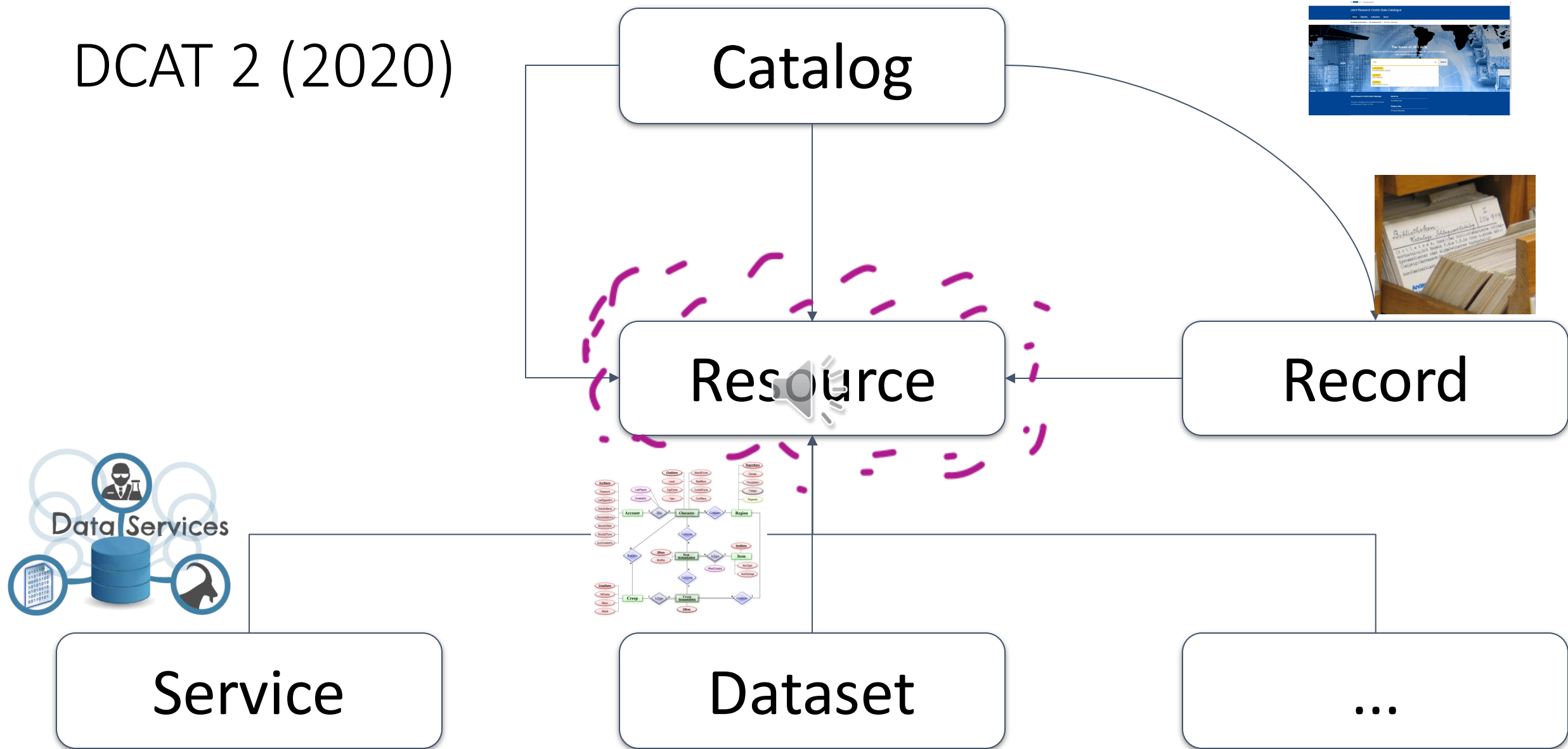
Dataset

Record

Distribution



DCAT 2 (2020)



DCAT – Principal Attributes

- Agents (creator, publisher, contact point)
- Spatial and temporal characteristics (coverage and resolution)
- Use and access conditions (rights, licenses, access conditions)



- Provenance
- Quality and conformance
- Qualified relationships

DCAT 3 (due at end of 2021)

- Dataset series
- Versioning

DCAT adoption

- Used across Europe as domain-independent metadata interchange format – e.g. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/201-0>
- Definition of national and domain-specific profiles (geospatial, statistical, transport, earth sciences) e.g. <https://data.overheid.nl/en/ondersteuning/open-data/dcat>, <https://inspire.ec.europa.eu/good-practice/geodcat-ap>
- The result is an active community providing implementation experiences that help consolidate the vocabulary, and adapt it to new needs

Status

- DCAT version 3 is still in development
- The GitHub is active – so follow discussions and contribute there.
- The sections indicated as 'normative' in the specification document (in HTML) are the only 'normative' elements of the vocabulary. Other sections and other representations are provided as informative).
- DCAT is a “spine” for other developments. It is an interoperability standard. We expect application profiles to be developed. See EC SEMICeu examples. These also include SHACL shapes for DCAT-AP.
- <https://w3c.github.io/dxwg/dcat/>
- <https://github.com/w3c/dxwg/tree/gh-pages/dcat>
- <https://github.com/SEMICeu/DCAT-AP>

DCAT: Next steps

- Evergreen standard - DCAT will be regularly updated to address users' requirements
- But we want to go beyond maintenance, and investigate how to make DCAT deal with new challenges in data management and discovery
- You input would be more than welcome!



<https://github.com/w3c/dxwg/issues>