GGBN – Strategies for standardized exchange of genetic ressources on a global scale

GGBN - Strategien für standardisierten Austausch genetischer Ressourcen auf globaler Ebene

Gabi Dröge

Botanic Garden and Botanical Museum Berlin

Global Genome Biodiversity Network

- **Network for non-human biobanks** (e.g. DNA, tissue)
- Founded in 2011
- Precursor project DNA Bank Network founded in 2007 providing virtual infrastructure
- General secretariat: Smithsonian Institution
- Technical secretariat: Botanic Garden and Botanical Museum Berlin
- http://www.ggbn.org

Goals

- Data standard for sharing tissue and DNA information
- Portal to make biobank sample data available
- Institutional directory
- Knowledge platform

Goals

- Best practices related to management and stewardship of genomic samples
- Recruit partners with different regional and taxonomic focus
- Identify gaps in GGBN collections

Audiences

- Biorepositories (contributors)
- Organizations with living or preserved specimens (contributors)
- Researchers (users)

Framework

- Memorandum of Cooperation
- Terms of Reference
- Business Model will come into force 06/2016
- Core and Associate members made a commitment to become financial or in-kind contributors

- Current funding by:

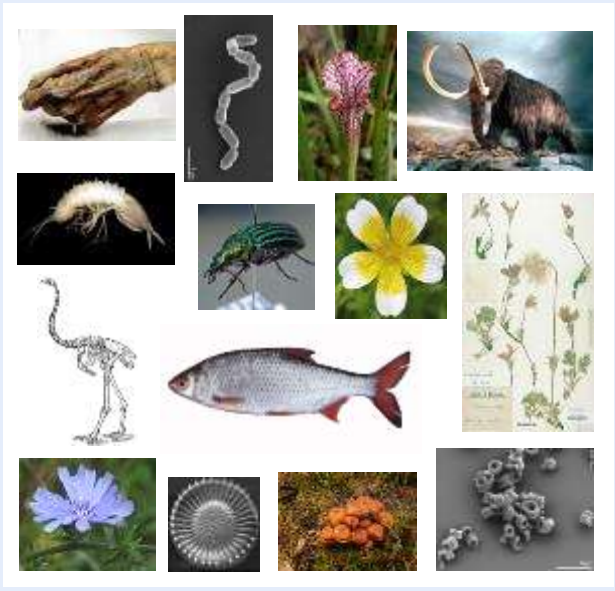
DFG SYNTHESYS Synthesis of systematic resources

Today: 40 GGBN members worldwide

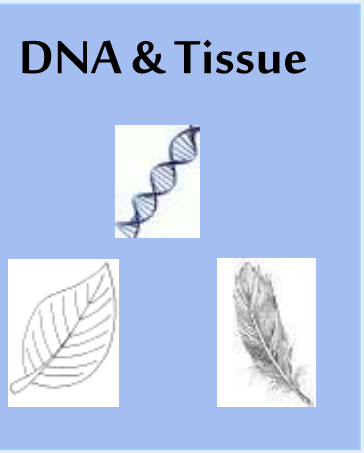
GGBN Data Portal

Data Portal Architecture

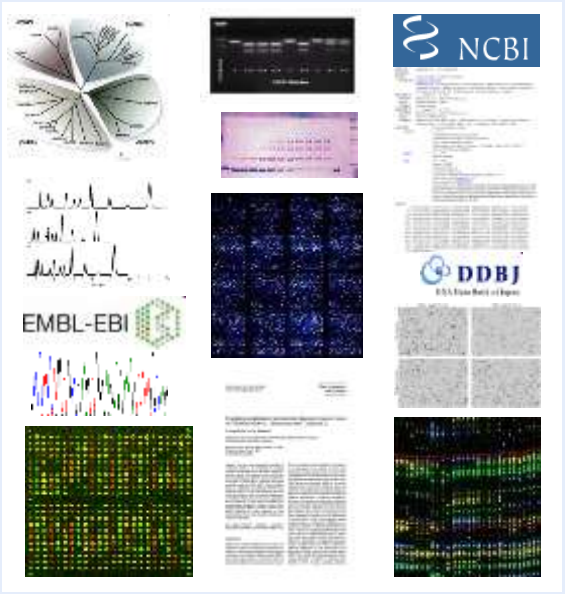
Source material / specimens



DNA & Tissue



Molecular analysis data



GBIF

GGBN

EMBL-EBI

DDBJ

NCBI

Global Biodiversity Information Facility (GBIF)

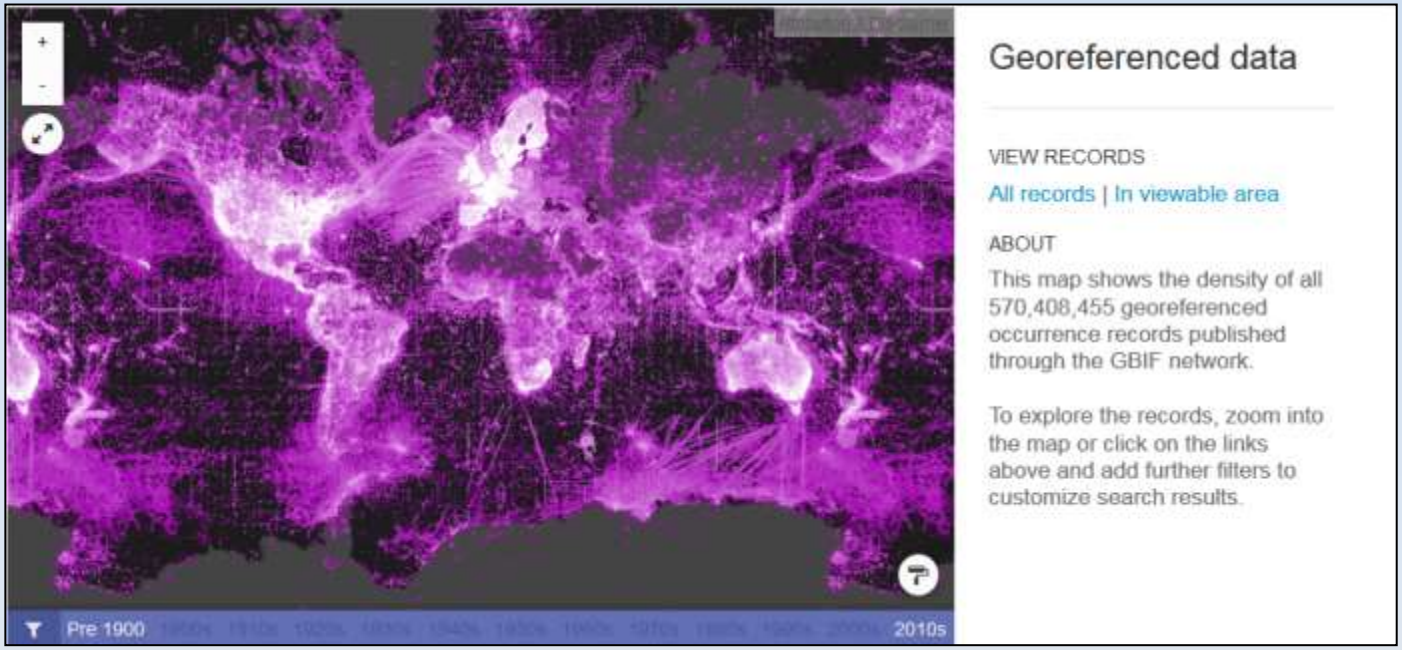
- GBIF's mission is to make the world's biodiversity data freely and universally available via the Internet.

- As a megascience initiative, GBIF aims to provide an essential global informatics infrastructure for biodiversity research and applications worldwide.

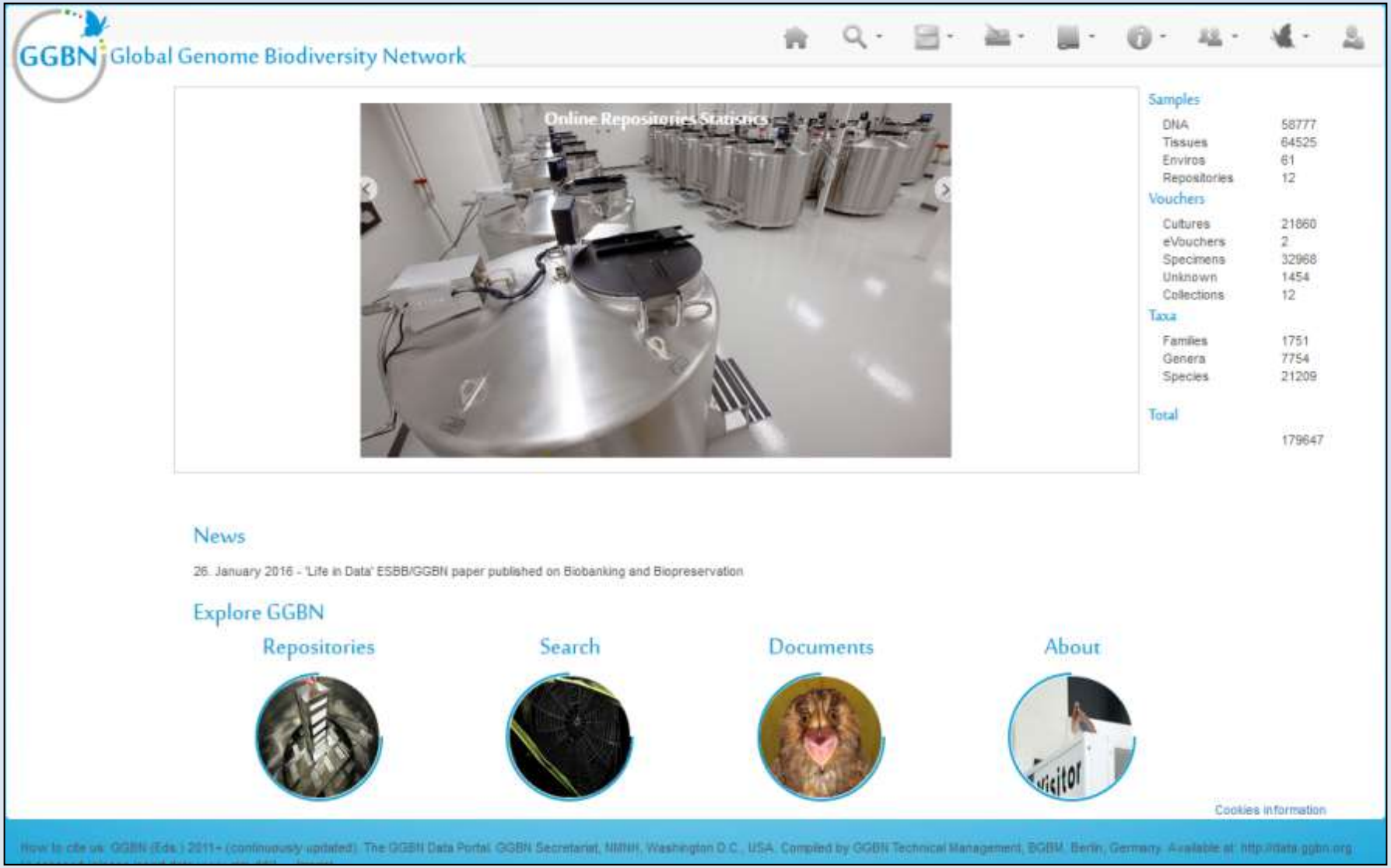
- http://www.gbif.org

GBIF - A data treasure of inestimable value

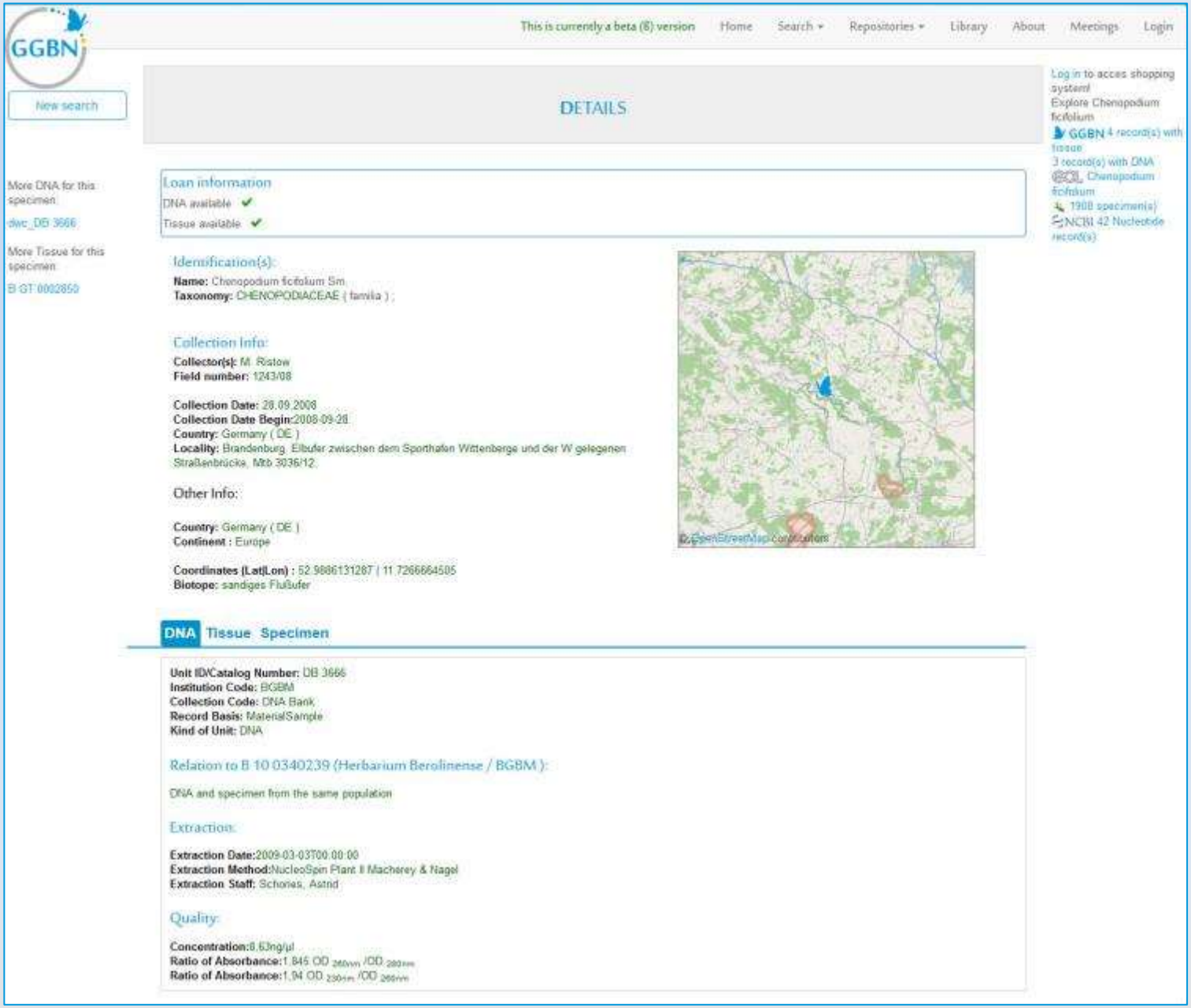
- > 797 data providers, including most of major natural history collections
- > 648.000.000 records



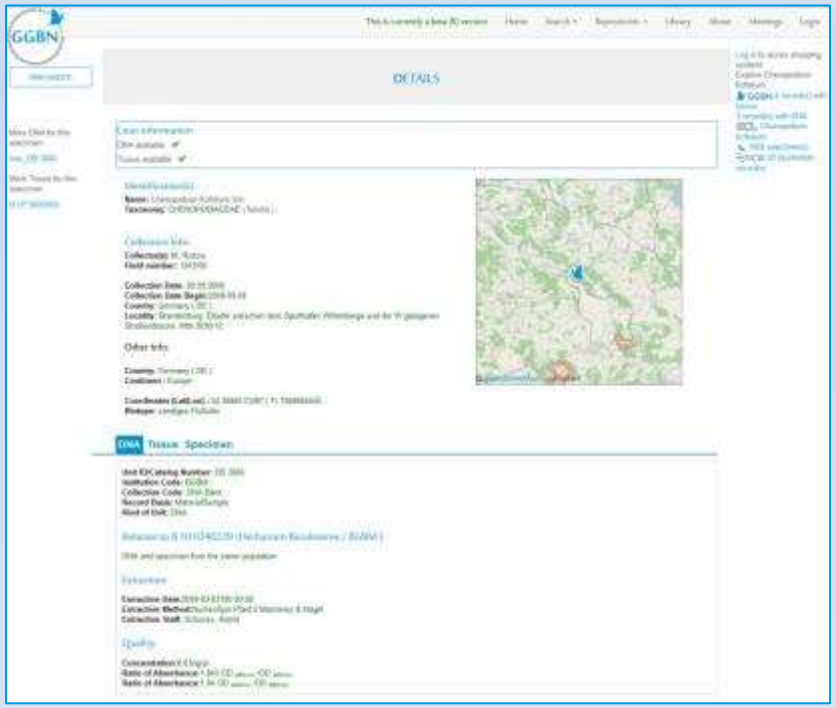
GGBN Data Portal



Portal: Aggregate data from multiple sources



Portal: Aggregate data from multiple sources



Getting live counts from other biodiversity portals for each record via web services

Explore *Chenopodium ficifolium*

GBIF 1908 specimens

NCBI 42 nucleotide sequences

eol Encyclopedia of Life taxon page

GGBN 3 DNA samples

4 tissue samples

CITES @ GGBN

This is currently a beta (ß) version Hon

DETAILS

Loan information

DNA available ✔

⚠ This species or genus was found in the CITES list (UNEP-WCMC (Comps.) 2015. More details here

Identification(s):

Name: Lophophora williamsii (Lem. ex Salm-Dyck) J.M.Coult.
Date: DateText: 17.06.2002
Identifier: B. E. Leuenberger
Taxonomy: CACTACEAE (familia) ;

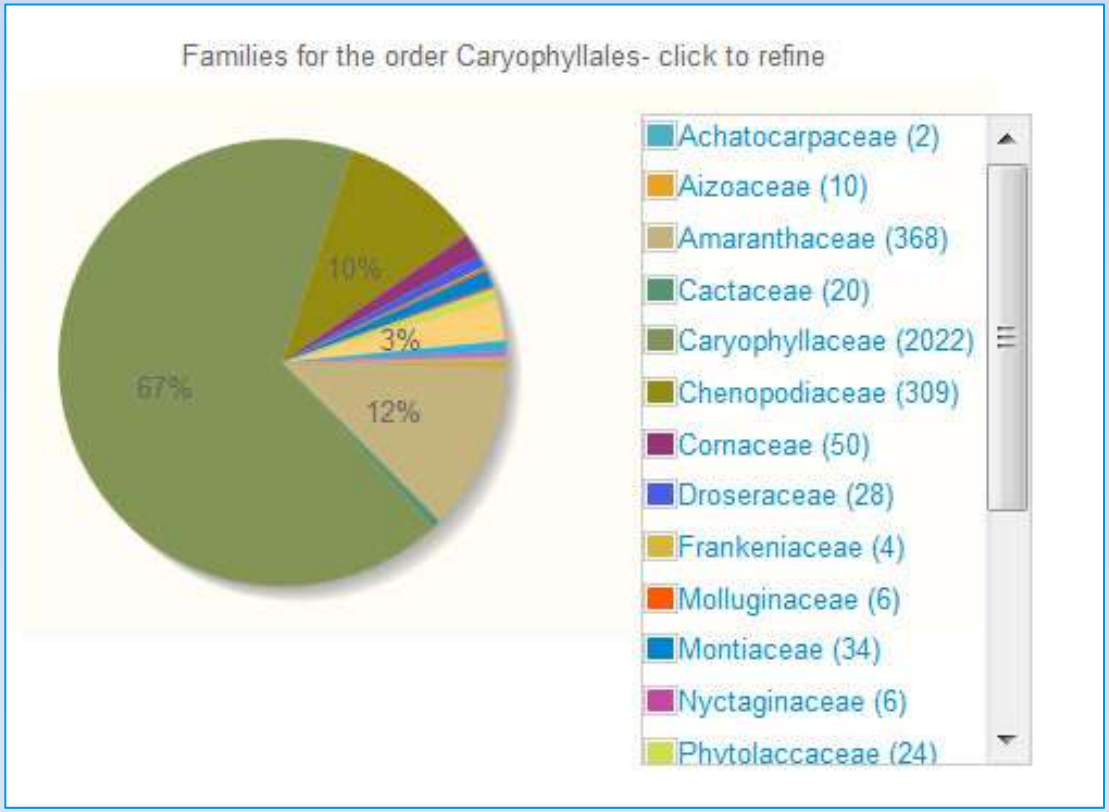
Warning and request for CITES registration number when ordering the sample both for curator and user.

Collection Info:

Collector(s): M. Cubr
Field number: 33663

Collection Date: 04.05.1995
Collection Date Begin:1995-05-04
Locality: Herkunft: BG München 1970; cult. Hort. bot. Berol.

Portal: Statistics



Example: Samples from Caryophyllales

Above: all records (DNA, tissue, specimens)

Right: DNA / tissue samples

⊞ Buxales (1 / 2)
⊞ Canellales (1 / 1)
⊞ Capparales (34 / 81)
⊟ Caryophyllales (290 / 1197)
⊞ Achatocarpaceae (0 / 0)
⊞ Aizoaceae (2 / 4)
⊞ Amaranthaceae (27 / 60)
⊞ Anacampserotaceae (0 / 0)
⊞ Ancistrocladaceae (0 / 0)
⊞ Asteropeiaceae (0 / 0)
⊞ Barbeuiaceae (0 / 0)
⊞ Basellaceae (0 / 0)
⊞ Cactaceae (6 / 11)
⊞ Caryophyllaceae (201 / 1026)
⊞ Cornaceae (11 / 26)
⊞ Didiereaceae (0 / 0)
⊞ Dioncophyllaceae (0 / 0)
⊞ Droseraceae (10 / 14)
⊞ Drosophyllaceae (0 / 0)
⊞ Frankeniaceae (1 / 2)
⊞ Gisekiaceae (0 / 0)
⊞ Halophytaceae (0 / 0)
⊞ Kewaceae (0 / 0)
⊞ Limeaceae (0 / 0)
⊞ Lophiocarpaceae (0 / 0)
⊞ Macarthuriaceae (0 / 0)
⊞ Microteaceae (0 / 0)
⊞ Molluginaceae (2 / 4)
⊞ Montiaceae (3 / 4)
⊞ Nepenthaceae (0 / 0)
⊞ Nyctaginaceae (2 / 2)
⊞ Physenaceae (0 / 0)
⊞ Phytolaccaceae (4 / 10)
⊞ Plumbaginaceae (18 / 30)
⊞ Portulacaceae (1 / 2)
⊞ Rhabdodendraceae (0 / 0)
⊞ Sarcobataceae (0 / 0)
⊞ Simmondsiaceae (1 / 1)
⊞ Stegnospermataceae (0 / 0)
⊞ Talinaceae (1 / 1)
⊞ Celastrales (14 / 27)

White Paper on Data Portal

Droege et al. 2014: NAR 42 (D1): D607-D612.

Nucleic Acids Research Advance Access published November 19, 2013

Nucleic Acids Research, 2013, 1–6
doi:10.1093/nar/gkt928

The Global Genome Biodiversity Network (GGBN) Data Portal

Gabriele Droege[1,*], Katharine Barker[2], Jonas J. Astrin[3], Paul Bartels[4], Carol Butler[2], David Cantrill[5], Jonathan Coddington[2], Félix Forest[6], Birgit Gemeinholzer[7], Donald Hobern[8], Jacqueline Mackenzie-Dodds[9], Éamonn Ó Tuama[8], Gitte Petersen[10], Oris Sanjur[11], David Schindel[2] and Ole Seberg[10]

[1]Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin 14195, Germany, [2]National Museum of Natural History Smithsonian Institution, Washington DC 20013, USA, [3]Zoological Research Museum Alexander Koenig, Bonn 53113, Germany, [4]Wildlife & Environment Society of South Africa, Pretoria 0001, South Africa, [5]National Herbarium of Victoria, Royal Botanic Gardens Melbourne, South Yarra, VIC 3141, Australia, [6]Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK, [7]Systematic Botany, Justus-Liebig-Universität, Giessen 35392, Germany, [8]Global Biodiversity Information Facility (GBIF), Copenhagen Ø DK-2100, Denmark, [9]Department of Zoology, The Natural History Museum, London SW7 5BD, UK, [10]Natural History Museum of Denmark, Copenhagen K DK-1307, Denmark and [11]Smithsonian Tropical Research Institute, Balboa Ancon, Unit 0948, Panama

Received August 15, 2013; Revised September 16, 2013; Accepted September 20, 2013

ABSTRACT

The Global Genome Biodiversity Network (GGBN) was formed in 2011 with the principal aim of making high-quality well-documented and vouchered collections that store DNA or tissue samples of biodiversity, discoverable for research through a networked community of biodiversity repositories. This is achieved through the GGBN Data Portal (http://data.ggbn.org), which links globally distributed databases and bridges the gap between biodiversity repositories, sequence databases and research results. Advances in DNA extraction techniques combined with next-generation sequencing technologies provide new tools for genome sequencing. Many ambitious genome sequencing projects with the potential to revolutionize biodiversity research consider access to adequate samples to be a major bottleneck in their workflow. This is linked not only to accelerating biodiversity loss and demands to improve conservation efforts but also to a lack of standardized methods for providing access to genomic samples. Biodiversity biobank-holding institutions urgently need to set a standard of collaboration towards excellence in collections stewardship, information access and sharing and responsible and ethical use of such collections. GGBN meets these needs by enabling and supporting accessibility and the efficient coordinated expansion of biodiversity biobanks worldwide.

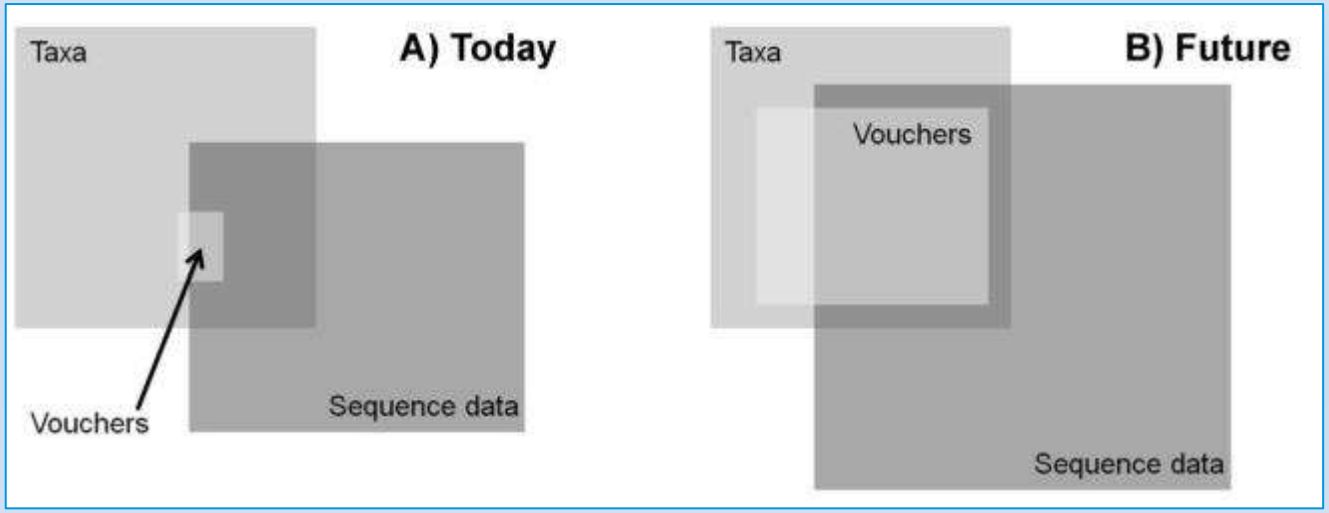
INTRODUCTION

Genome sequencing for biodiversity analysis is at the forefront of innovation and discovery due to technological advances and the sequencing of whole genomes in the last 10 years. Information generated from biodiversity genomics will revolutionize our approach to taxonomy, phylogeny, conservation, ecological monitoring, wildlife management, agriculture, drug development, zoonotic disease forecasting and even aspects of national security. Consequently, the demand is rapidly increasing for professionally preserved, managed and documented samples that yield high-molecular weight DNA and RNA from throughout the tree of life [e.g. (1,2)]. Many ambitious projects with the potential to revolutionize biodiversity research are finding access to adequate samples needed for genome sequencing to be a major bottleneck in their workflow. Examples of these projects include the Ten Thousand Vertebrate Genomes Project (Genome10 K),

*To whom correspondence should be addressed. Tel: +49 30 838 50 139; Fax: +49 30 838 50 186; Email: g.droege@bgbm.org
Present address:
Gabriele Droege, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin, 14195, Germany.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Vouchers, Traceability, Deposition



Source: Droege et al. 2014

1. Every biodiversity biorepository is welcome to join GGBN.
2. Deposit your samples, vouchers and data in a GGBN collection if you don't have a DNA or tissue bank.
3. GGBN provides a virtual and physical infrastructure to make your research traceable for the future.

Portal: Pre-order samples

- Order DNA or tissue samples through GGBN portal
- Download of sample information
- Request forwarded to GGBN member holding the sample(s)
  - Institution responsible for all further steps
    - Checking availability and loaning conditions
    - Provide price offer to scientist
    - Request signing of Material Transfer Agreement
    - Shipping samples
- Citation guidelines for samples coming soon (available at ggbn.org)

Developing Best Practices Towards a Network of Trusted Biodiversity Biobanks

Developing Best Practices

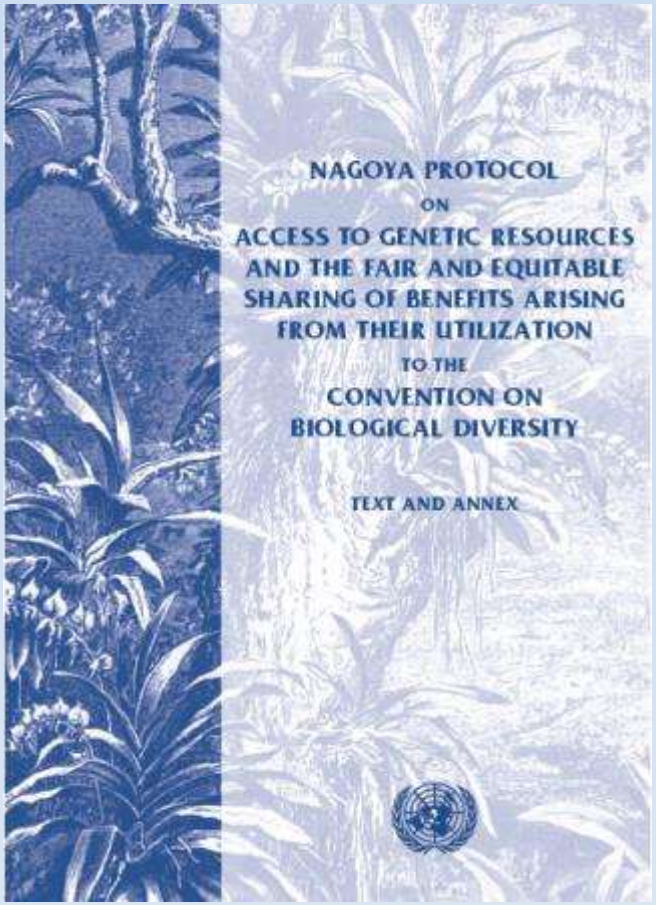
- Recommendation for Biodiversity Biorepositories
  - **Collaboration with ISBER***
  - April 2013, Submitted to ISBER
  - To be included into next version of ISBER Best Practices
- GGBN Document Library
  - **Collaboration with ESBB****
  - June 2016, beta-release for member use
  - Knowledge platform for non-human biobanking

*** International Society for Biological and Environmental Repositories**

**** European, Middle Eastern and African Society for Biopreservation and Biobanking**

Developing Best Practices

- Access and Benefit Sharing (ABS) Compliance
  - October 2014, Documentation available for member use
    - Material Transfer Agreements
    - Code of Conduct
    - Statement of Use of Genomic Material
  - **Collaboration with CETAF***
  - Provide trusted and transparent access to genomic samples for users and contributors through an ABS framework



*** Consortium of European Taxonomic Facilities**

GGBN documents

- MTA for provision of material
  - with no change in ownership
  - with change in ownership
- MTA for receipt of material with change in ownership
- Best Practice for Access and Benefit Sharing
- Code of Conduct

- http://wiki.ggbn.org/ggbn/Downloads

GGBN documents

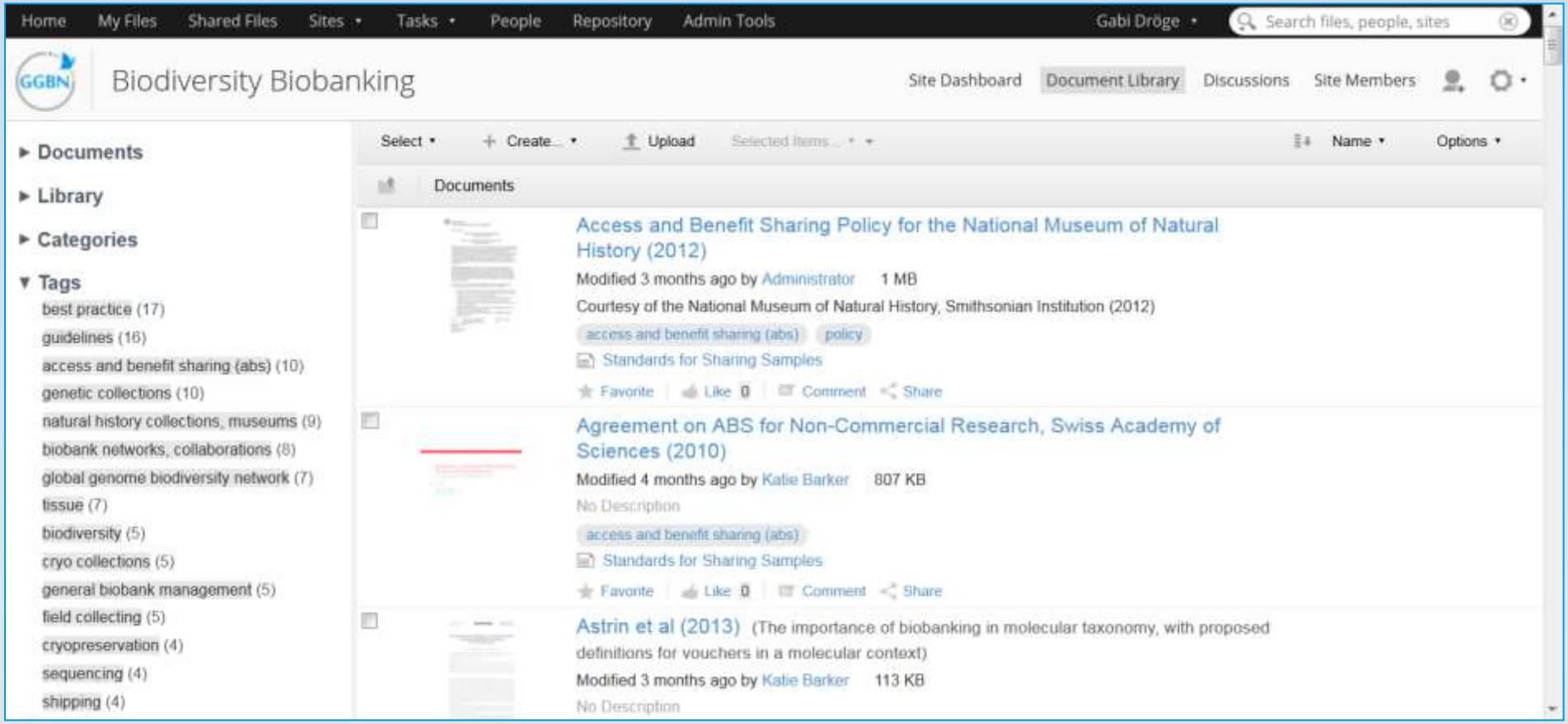
→To be used for DNA and tissue samples, but also other genetic ressources

→To be used by GGBN members and third parties as templates

GGBN Document Library

Alfresco Community

- full text search, tagging, versioning, sharing
- release prototype 06/2016



Data Portal

Biodiversity is challenging

One species – three use cases

eVoucher (image)

Blood sample

Genomic DNA sample

gDNA

A few sequences, one species

No voucher

Fecal sample

Environmental DNA sample

eDNA

Thousands of sequences, dozens of species

Ancient specimen as voucher

Bone sample

Ancient DNA sample

aDNA

A few fragmented sequences, one species

GGBN Data Standard

http://terms.tdwg.org/wiki/GGBN_Data_Standard

GGBN Data Standard

The Global Genome Biodiversity Network (GGBN) is a global network of well-managed collections of genomic tissue samples from across the Tree of Life, benefiting society through biodiversity research, development and conservation. This network will foster collaborations among repositories of molecular biodiversity in order to ensure quality standards, improve best practices, secure interoperability, and harmonize exchange of material in accordance with national and international legislation and conventions.

The GGBN Data Standard is a set of vocabularies designed to represent tissue, DNA or RNA samples associated to voucher specimens, tissue samples and collections. *Contributors: Gabriele Droege, Birgit Gemeinholzer, Holger Zetzsche, Astrid Schories, Katharine Barker, Walter G. Berendsohn, Sean Brady, E. Margaret Casey, Jonathan Coddington, John Deck, Anton Güntsch, Jörg Holetschek, Patricia Kelbert, Hans-Peter Klenk, Renzo Kottmann, Christopher Lewis, Jacqueline Mackenzie-Dodds, Christopher Meyer, Thomas Orrell, Ole Seberg, Jamie Whitacre, Pelin Yilmaz, Enrique Arbeláez Cortés, Boyke Bunk, René Dekker, Sonya Dyhrman, Elisabeth Haring, Thomas Knebelsberger, Jon Norenburg, Michael Raupach, Thomas von Rintelen, Larissa Smirnova, Carola Söhngen, Sun Ying, Lee A. Weigt, Kenneth Wurdack, Elizabeth Zimmer, Xin Zhou.*

Contents [hide]

GGBN Data Standard

Concept definitions in GGBN Permit Vocabulary

Concept Name:	**ggbn:permitStatus**
Normative URI	http://rs.tdwg.org/dwc/terms/disposition
Label	Permit Status
Definition	Information about the presence, absence or other basic status of permits associated with the sample(s).
Defined By	
	Required: Yes — **Repeatable:** Yes
Examples	Permit available, Permit not required, Permit not available, Unknown **Material collected after 2014-10-12 cannot be in "Unknown" permit status!**
Notes:	Predefined vocabulary, See permitStatus vocabulary; entry of data is mandatory
Concept Name:	**ggbn:permitStatusQualifier**
Normative URI	http://rs.tdwg.org/dwc/terms/disposition
Label	Permit Status Qualifier
Definition	Description of why a certain permit was not required or why Permit Status is unknown
Defined By	
	Required: No — **Repeatable:** Yes
Examples	"no national requirement for a permit at date of access", "officially authorized illegal holder", "collected on private land", "pre-Nagoya"
Concept Name:	**ggbn:permitText**
Normative URI	http://rs.tdwg.org/dwc/terms/disposition

GGBN Data Standard

- http://terms.tdwg.org/wiki/GGBN_Data_Standard
- Based on ABCDDNA
- Is meant to be used with ABCD or DwC -> all occurrence terms are excluded (geography, scientificname etc.)
- White paper work in progress
- Include elements of other standards
- Can handle or be combined with SPREC, MIxS and BRISQ
- **Collaboration with GBIF, Genomics Standards Consortium, ESBB**

GGBN Data Standard

Page | Discussion | Read | View form | View source | View history | Search

ggbn:preparationMaterials

Preparation Materials: Materials and chemicals used in the preparation of the specimen, tissue, DNA or RNA sample

Example(s): for DNA: DNeasy blood and tissue kit, CTAB;

Constraints in the context of GGBN Data Standard:

- cardinality 0..n (=optional multiple occurrence)

Scheme: GGBN Data Standard
Collection: GGBN Preparation Vocabulary

Preparation Materials

- *URI:* http://data.ggbn.org/schemas/ggbn/terms/preparationMaterials
- *skos: has exact match* mixs:nucl_acid_ext

RDF feed | Browse properties | SMW-prop.

Search for values

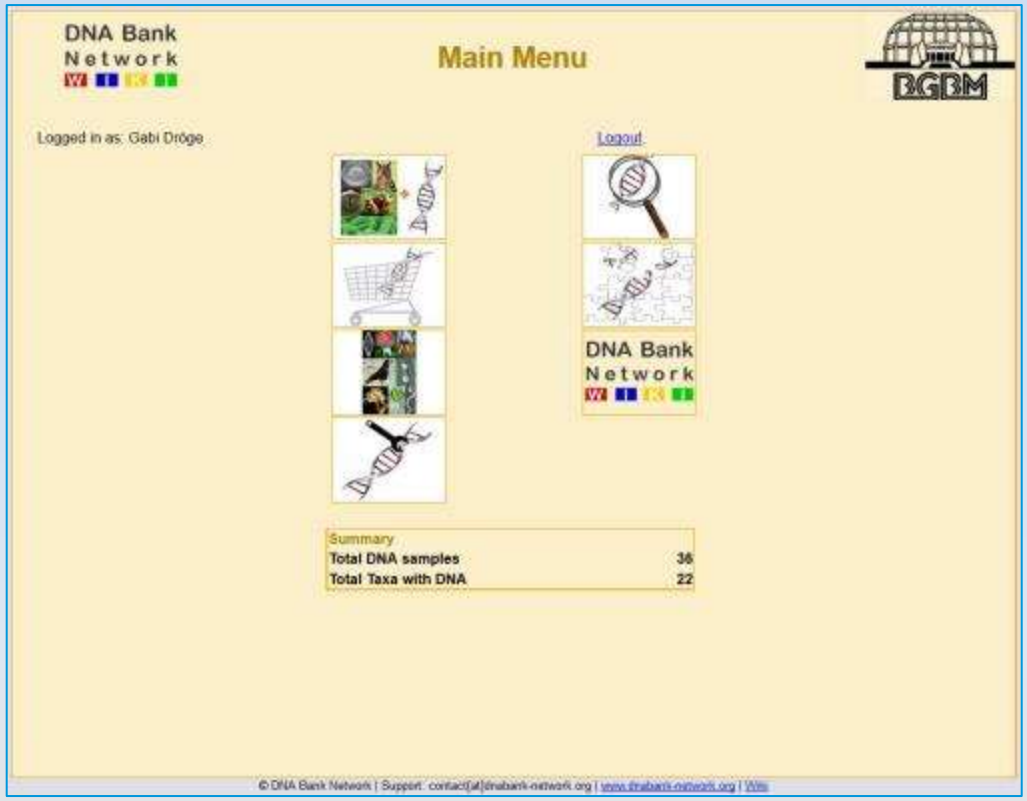
Classes: Property | Concept | GGBN Data Standard

Implementing the GGBN Data Standard

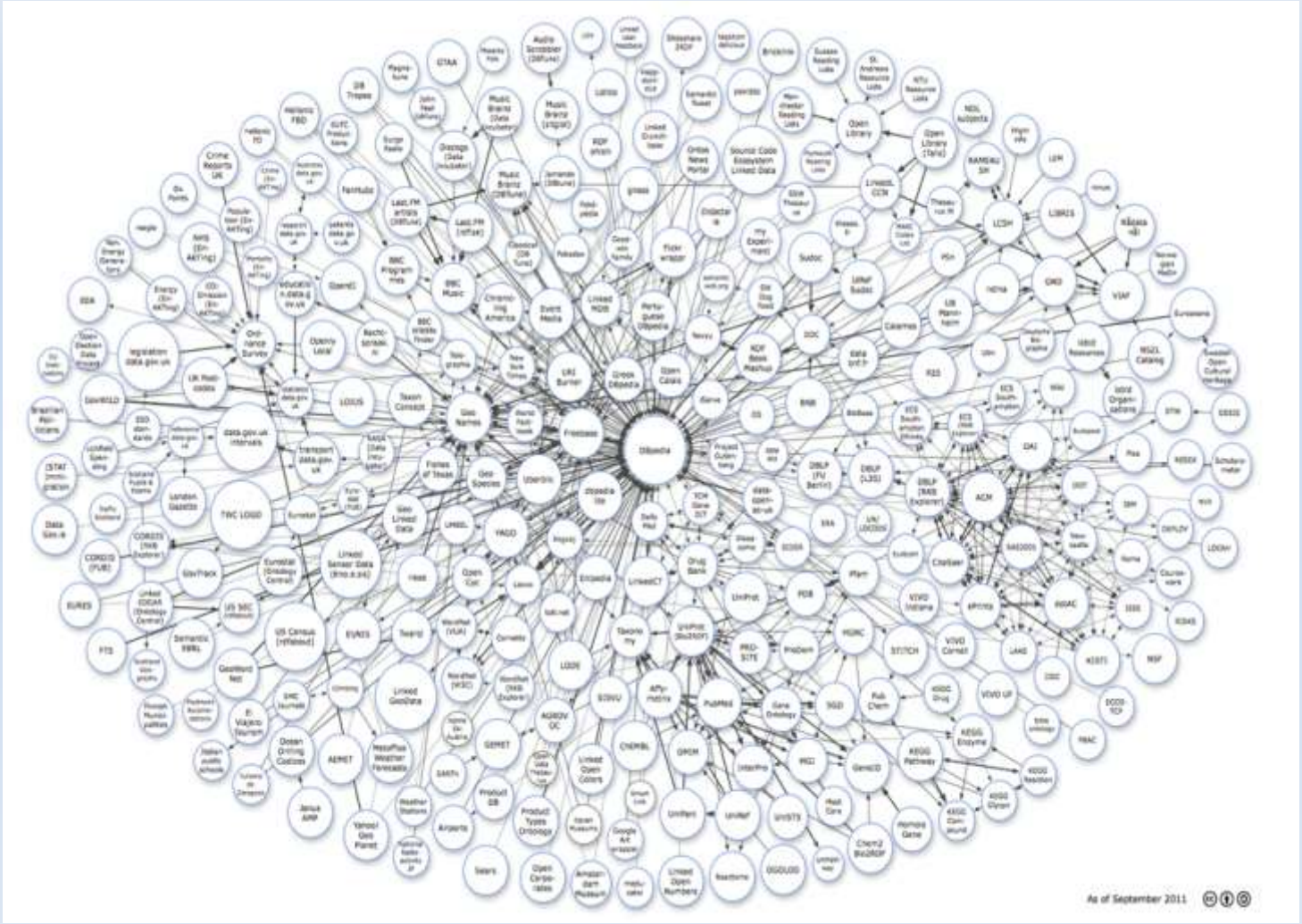
- Done for GGBN Data Portal
- To be done: databases of GGBN members must be updated
- Challenging:
  - different databases and software systems in use
  - many partners don‘t even have a database

DNA Module

- Open source software developed by BGBM/DNA Bank Network for management of DNA and tissue collections
- 100% in compliance with GBIF and GGBN



Future: Linked Data and Stable Identifiers



GGBN collaborates with other stakeholders and across domains to reach this goal.

Second GGBN International Conference

21.-24. June 2016 Berlin

BG BM | Botanischer Garten & Botanisches Museum Berlin

museum für naturkunde berlin

Second GGBN International Conference

- Interest Group Meeting: Access and Benefit Sharing
- Symposium: Best practice for data and collections management
- Workshop: ABS in natural history collections – implementation and practical management
- http://meetings.ggbn.org

- Conference held in parallel with **SPNHC*** conference
- http://www.spnhc2016.berlin

*** Society for the Preservation of Natural History Collections**

Thank you

CC-BY-SA 3.0 G.Droege 2015-06-05 Phoenix, AZ

http://www.ggbn.org

info@ggbn.org

GGBN Interim Executive Committee

GGBN Members

GGBN Collaborators

GGBN Task Forces

DFG

SYNTHESYS

ISBER & ESBB