BIG4: Biosystematics, informatics and genomics of the big 4 insect groups- training tomorrow's researchers and entrepreneurs

Kick-Off Meeting

14-18 September 2015

Copenhagen, Denmark

DINA: An Open Source Collection Management System

Fredrik Ronquist

Dept. Bioinformatics and Genetics

Swedish Museum of Natural History

Naturhistoriska
riksmuseet

Collection Management Systems

Institutional Choices (forced on staff…):

1. Develop your own system in-house

2. Acquire a commercial system

3. Partner with other institutions in distributed open-source development

The Case For Open Source

- **Market considerations.** Professional collection management systems not viable commercial products in a pluralistic market.
- **Long-term stability**. An open-source software solution developed by institutions with long-term focus will be more stable than a commercial solution.
- **Flexibility**. A distributed open-source system must by necessity conform to a modular design based on open API:s. This favors flexibility and adaptability in a way that a commercial product will not.
- **Cost effectiveness**. Although some overhead is associated with distributed development, more development teams involved in the effort will result in a lower cost to the individual institution compared to in-house or commercial solutions.
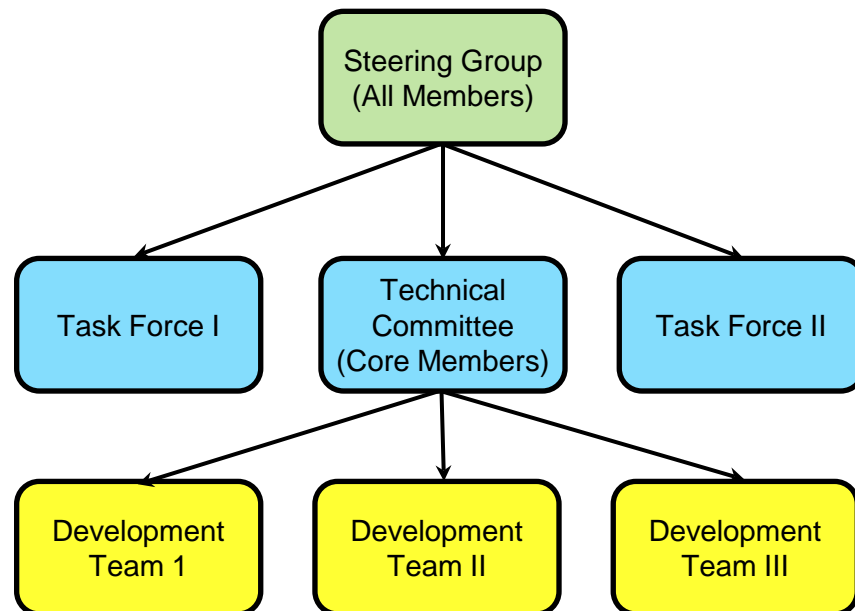
The Case For Open Source (cont'd)

- **Opt-in opt-out scheme**. Institutions can participate in the development when they have resources to do so, and can opt out when they do not. At any single point in time, it should be feasible to have enough institutions involved for development to move forward at an acceptable pace.

- **Community Control**. A distributed open-source solution means that the community retains control over both the information standards and the system architecture and web service/API designs.

- **Egalitarian**. A professional open-source collection management system offers a better way for developing countries to catch up than any commercial product.

- **Stable marketplace for extensions and services**. A community-supported de-facto standard for collection management systems architecture will ensure that there is a stable market for various plugins, extensions and services based on the system.

DINA Consortium

(Digital Information system for NAtural history data)

- **Core Member**. Required contribution 1.0 FTE to the project, of which at least 0.5 to the development effort. Voting member of the DINA Technical Committee (TC), which controls deliverables and deadlines for the 1.0 FTE contribution.

- **Associate Member**. No contribution requirements. Non-voting member of the Steering Group.

```
                    Steering Group
                    (All Members)
                          |
        ┌─────────────────┼─────────────────┐
        ↓                 ↓                 ↓
   Task Force I       Technical         Task Force II
                      Committee
                      (Core Members)
                          |
        ┌─────────────────┼─────────────────┐
        ↓                 ↓                 ↓
   Development       Development       Development
   Team 1            Team II           Team III
```

DINA Consortium

- Core Members
  - Agriculture and Agri-Food Canada, Ottawa
  - Estonia (University of Tartu)
  - Denmark (University of Copenhagen)
  - Sweden (Swedish Museum of Natural History)
- Associate Members
  - Museum für Naturkunde, Berlin
  - Royal Botanic Garden, Edinburgh
- Open to Additional Members
  - Memorandum of Cooperation and more information at http://dina-project.net

Lessons Learned

- **Commitment**. Formalization of the collaboration and a good governance model is essential.

- **Patience**. It may take an institution with long-term perspective several years from a decision to join the consortium to actively contributing to the development.

- **Respect**. Different teams come with different backgrounds, different skill sets, and different external pressures. Striking the right balance between the cathedral (centrally controlled) and the bazaar (locally controlled) approach to collaborative development is crucial.

- **Trust**. A team needs to trust the other teams in the consortium to deliver according to agreements, so that consortium membership pays off.
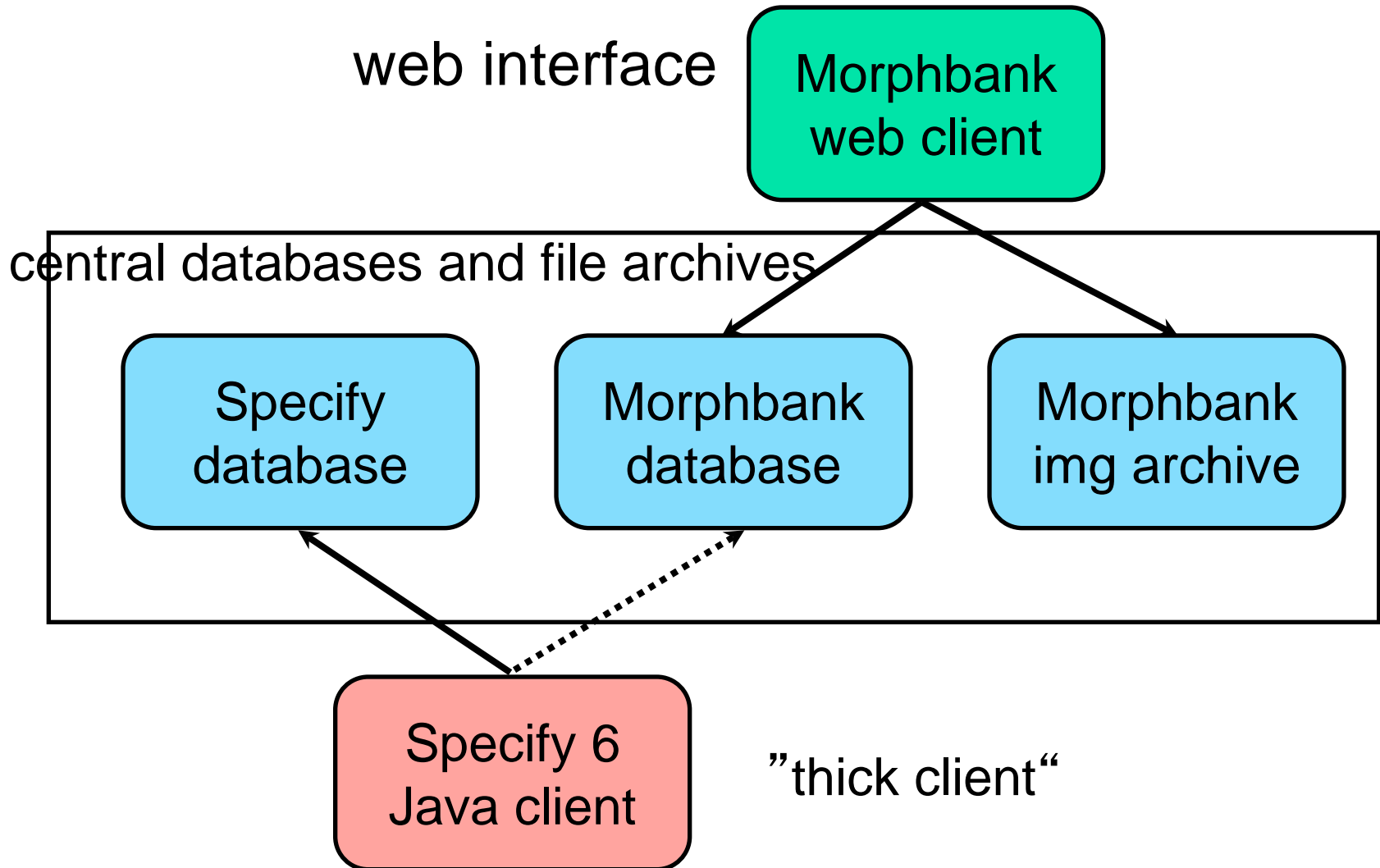
DINA Versions

- **DINA Light** ("Specify")

  - Based largely on Specify 6 and the Specify data model, combined with new API:s and web clients (collection web portal, biological survey client, species pages, DNA barcode portal, loan request system)

  - Fully compatible with Specify 7

  - In production in Sweden since 2011. Currently includes many of the small Swedish collection databases (NRM entomology, geology; GNM entomology, SMTP) with several more on the way in (NRM zoology (part) and paleontology, GNM zoology and geology, EMU geology).

- **DINA Web**

  - Modular service-oriented architecture, based to a large extent on the Specify data model

  - DINA API guidelines and style guidelines adopted

  - Architectural road map, module overview and API blueprints under discussion

  - Core modules available in proto-DINA versions: collection web portal, species pages system, biological survey client, DNA barcode portal

  - Core modules under development: taxonomy module, collection manager, DNA sequence module, DINA data tool (batch uploading and editing)

DINA Light

web interface

Morphbank web client

central databases and file archives

Specify database

Morphbank database

Morphbank img archive

Specify 6 Java client

"thick client"

DINA Web

Collection Manager

Biodiversity survey client

Collection web portal

DNA barcode portal

Species pages

Specify database

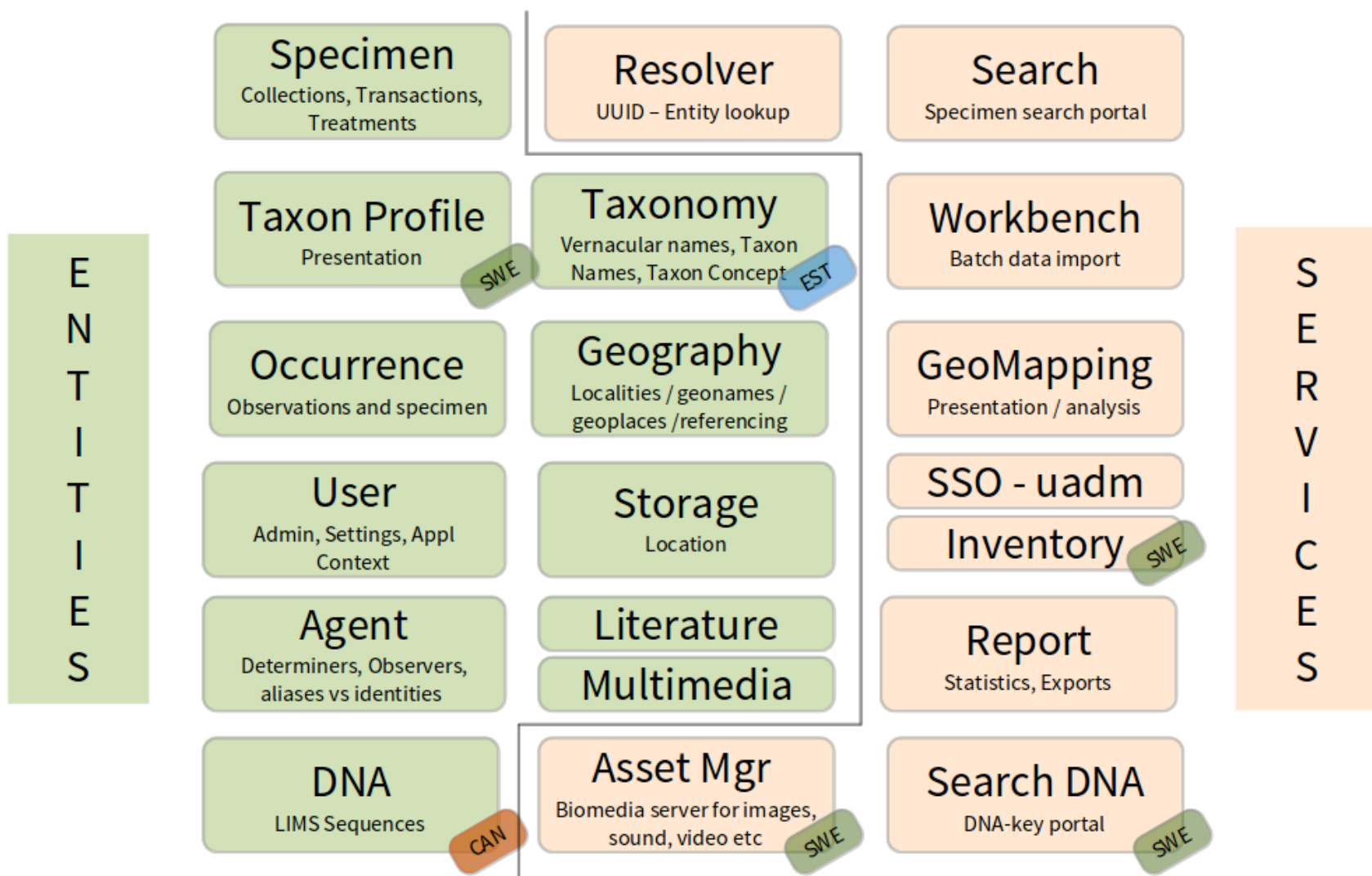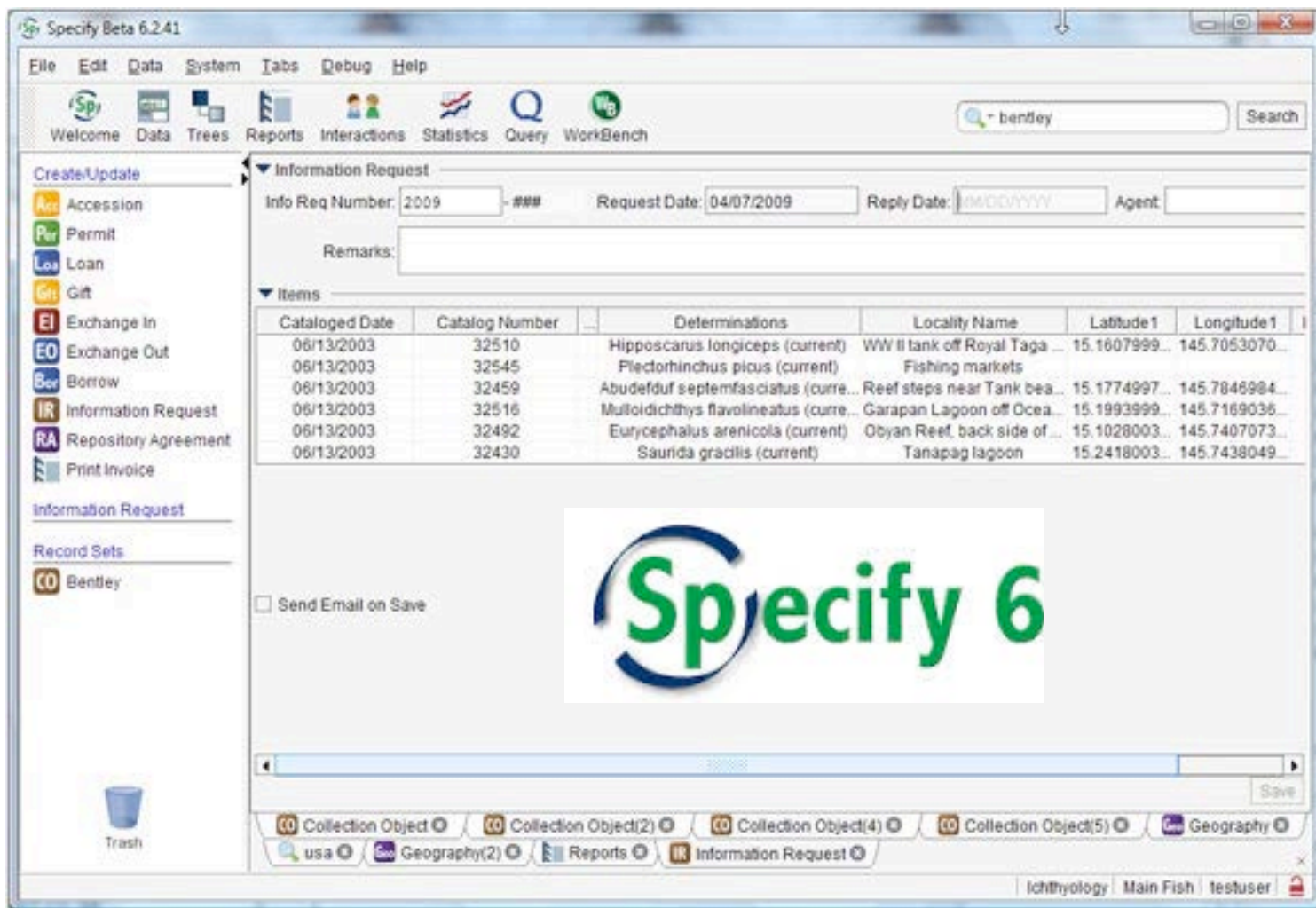Media metadata

Media files

BLAST DB

Taxon info

ENTITIES

Specimen
Collections, Transactions, Treatments

Taxon Profile
Presentation
SWE

Taxonomy
Vernacular names, Taxon Names, Taxon Concept
EST

Occurrence
Observations and specimen

Geography
Localities / geonames / geoplaces /referencing

User
Admin, Settings, Appl Context

Storage
Location

Agent
Determiners, Observers, aliases vs identities

Literature

Multimedia

DNA
LIMS Sequences
CAN

SERVICES

Resolver
UUID – Entity lookup

Search
Specimen search portal

Workbench
Batch data import

GeoMapping
Presentation / analysis

SSO - uadm

Inventory
SWE

Report
Statistics, Exports

Asset Mgr
Biomedia server for images, sound, video etc
SWE

Search DNA
DNA-key portal
SWE

Entities – what are the relationships between entities? Info model
Services – what are the functions and which entities are involved in each unit? API

File Edit Data System Tabs Debug Help

Welcome | Data | Trees | Reports | Interactions | Statistics | Query | WorkBench

[Search box: ▾ bentley] Search

Create/Update

Acc Accession
Per Permit
Loa Loan
Gft Gift
EI Exchange In
EO Exchange Out
Bor Borrow
IR Information Request
RA Repository Agreement
Print Invoice

Information Request

Record Sets

CO Bentley

▾ Information Request

Info Req Number: 2009 - ###   Request Date: 04/07/2009   Reply Date: MM/DD/YYYY   Agent:

Remarks:

▾ Items

Cataloged Date	Catalog Number	...	Determinations	Locality Name	Latitude 1	Longitude 1	
06/13/2003	32510		Hipposcarus longiceps (current)	WW II tank off Royal Taga ...	15.1607999...	145.7053070...	
06/13/2003	32545		Plectorhinchus picus (current)	Fishing markets			
06/13/2003	32459		Abudefduf septemfasciatus (curre...	Reef steps near Tank bea...	15.1774997...	145.7846984...	
06/13/2003	32516		Mulloidichthys flavolineatus (curre...	Garapan Lagoon off Ocea...	15.1993999...	145.7169036...	
06/13/2003	32492		Eurycephalus arenicola (current)	Obyan Reef, back side of ...	15.1028003...	145.7407073...	
06/13/2003	32430		Saurida gracilis (current)	Tanapag lagoon	15.2418003...	145.7438049...	

☐ Send Email on Save

Specify 6

Trash

Save

CO Collection Object ⊗ CO Collection Object(2) ⊗ CO Collection Object(4) ⊗ CO Collection Object(5) ⊗ Geo Geography ⊗
🔍 usa ⊗ Geo Geography(2) ⊗ Reports ⊗ IR Information Request ⊗

Ichthyology | Main Fish | testuser

Species Pages

Collection Web Portal

https 🔒 dev.plutof.ut.ee

Reader

PlutoF

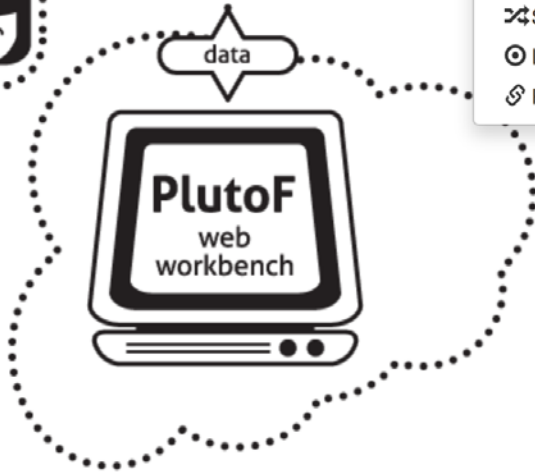🗎 Studies　　👤 Persons　　⛰ References　　🔡 Laboratory ▾　　⬛ Taxon Occurrences ▾　　🚩 Clipboard　　🔧 Settings　　↦ Logout

🍃 Specimens
👁 Observations
⤭ Sequences
◎ Living Culture
🔗 Reference-Based

data

PlutoF
web
workbench

Welcome to PlutoF

PlutoF provides cloud database and computing services for the taxonomical, ecological, phylogenetical, etc. research. The purpose of the platform is to provide synergy through common modules for the classifications, taxon names, analytical tools, etc.

It allows to address integrated questions in ecology and coevolution of taxa. Different types of the species occurrences, viz. preserved specimens, DNA sequences, human observations, references can be stored in PlutoF as well. PlutoF has no restrictions on taxon and geographic coverage and therefore can be used for the databasing interacting taxa.

It also includes collection management module. Few examples of the public web outputs from PlutoF are Estonian eBiodiversity (http://elurikkus.ut.ee), and molecular key for fungi (http://unite.ut.ee).

Feedback

NATARC

BalticDiversity

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

CENTRAL BALTIC
INTERREG IV A
PROGRAMME
2007–2013

UNIVERSITY OF TARTU
NATURAL HISTORY MUSEUM
1632

PlutoF

EU BON

Sequence Database

You have the following options available:

Collections

- Specimen Collections
- Specimens
- Specimen Export
- Specimen Loan
- Mixed Specimens
- Import Permits
- Samples

PCR

- Gene Regions
- PCR Primers
- PCR Profiles
- Product Inventory
- PCR Batches
- PCR Reactions

Sequencing

- Sequence Batches
- Sequences

Storage

- Storage Units
- Storage Container Types
- Storage Containers

Miscellaneous

- Tasks
- Projects

Agriculture and
Agri-Food Canada

Agriculture et
Agroalimentaire Canada

Two routes into DINA in Sweden

"DINA Light"

Data cleaning
Migration

Specify 6
Morphbank

Minor adjustments

Legacy
system

DINA Web

Adjustment of data models
Adjustment of UIs and functionality
Data cleaning and migration
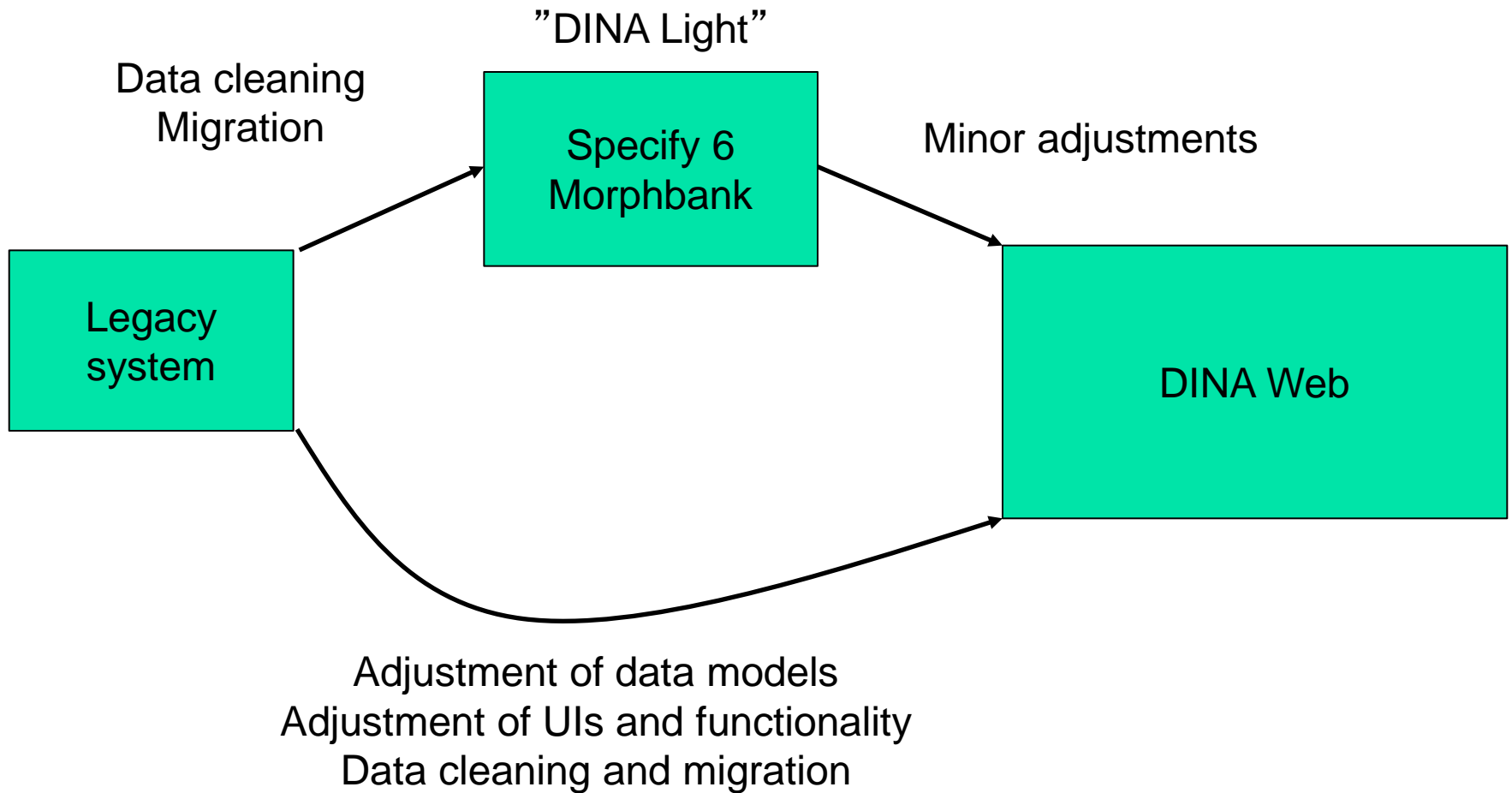
The DINA-IRIS team at NRM

Karin Karlsson

Kevin Holston

Markus Englund

Ida Li

Ingimar Erlingsson

Markus Skyttner

More DINA Info

- **DINA project wiki (http://dina-project.net)**
  - Project introduction
  - Steering committee and technical committee information, minutes of meetings etc
  - Status of the project in each of the participating institutions
- **DINA github repository (https://github.com/DINA-Web)**
  - DINA API guidelines and style guidelines
  - Module map, system overview
  - Code for DINA modules
- **DINA components in production in Sweden:**
  - **http://naturforskaren.se (species pages, in Swedish)**
  - **http://naturfynd.se (biodiversity survey client, requires login)**
  - **http://naturarv.se (collection web portal)**
  - **http://dna-key.se (DNA barcode portal)**
  - **https://www.dina-web.net/loan/ (loan request)**