pandata
Photon and Neutron
Data Infrastructure

Umbrella

EuroFEL
FREE ELECTRON LASERS OF EUROPE

European photon/neutron facilities
The User Umbrella System, Status and Future

Photon/Neutron Facilities and Authentication

TOC

- The community
- General characteristics
- IT requests
- Umbrella concept
- Authentication and authorization
- Roadmap
- Status and Outlook

The user community I

- ***Photon facilities***
  - ✓ Synchrotrons and Free Electron Lasers (FELs)
  - ✓ Produce light of highest brightness
  - ✓ Typical range from infra-red to Xrays
  - ✓ Size hundreds of meters
  - ✓ About 15 synchrotrons in EU
    - ESRF Grenoble
    - National facilities (DESY, HZB, PSI …)
- ***FELs, even 10^3 to 10^6 times brighter***
  - ✓ SLAC/Stanford, DESY/Hamburg, FEL/Spring-8/Japan, PSI/Villigen
  - ✓ Membrane proteins; microscopic movies of chemical reactions
- ***Neutron facilities***
  - ✓ Complementary
  - ✓ Similar user community
- ***Wide range of research areas***
  - ✓ Archaeology, chemistry, materials science, life sciences, physics …
- ***Small teams, visit for***
  - ✓ Few hours (structural biology) to
  - ✓ Few weeks (superconductivity, nano investigations)

The user community II

- ***In EU in the order of several 10'000 visiting users /y***
  - ✓ Large overbooking (≥3:1), low chance to be accepted
  - ✓ Important to minimize administrative load (Local user offices)
- ***On-site visits***
  - ✓ Short duration
  - ✓ In part spontaneous (keep that attraction)
- ***Decentralized structure (compare e.g. to CERN)***
  - ✓ Manifold research fields
  - ✓ Several facilities
- ***National character of facilities***
  - ✓ Report to national governments
- ***'Part-time' users***
  - ✓ E.g. structural biology: 10% of time
- ***But: These large facilities produce excellent results***
  - ✓ Standard tool in structural biology (e.g. genome research)
  - ✓ 2009 Nobel prizes in chemistry

What is the current situation?

- ***Small research groups***
  - ✓ Patchwork teams
  - ✓ In general low IT background
  - ✓ Visit for
    - Few hours (structural biology)
    - Few weeks (superconductivity, chemistry)
- ***Administration by local User Offices***
  - ✓ Tools: WUOs = Web-based User Offices
  - ✓ Users registered with local WUOs
  - ✓ Proposals as ordering elements
- ***No official cross-facility information exchange***
  - ✓ Competition among users
  - ✓ Competition among facilities
- ***Limited amount of data (Gbyte)***
  - ✓ Hard-disk in trouser pocket

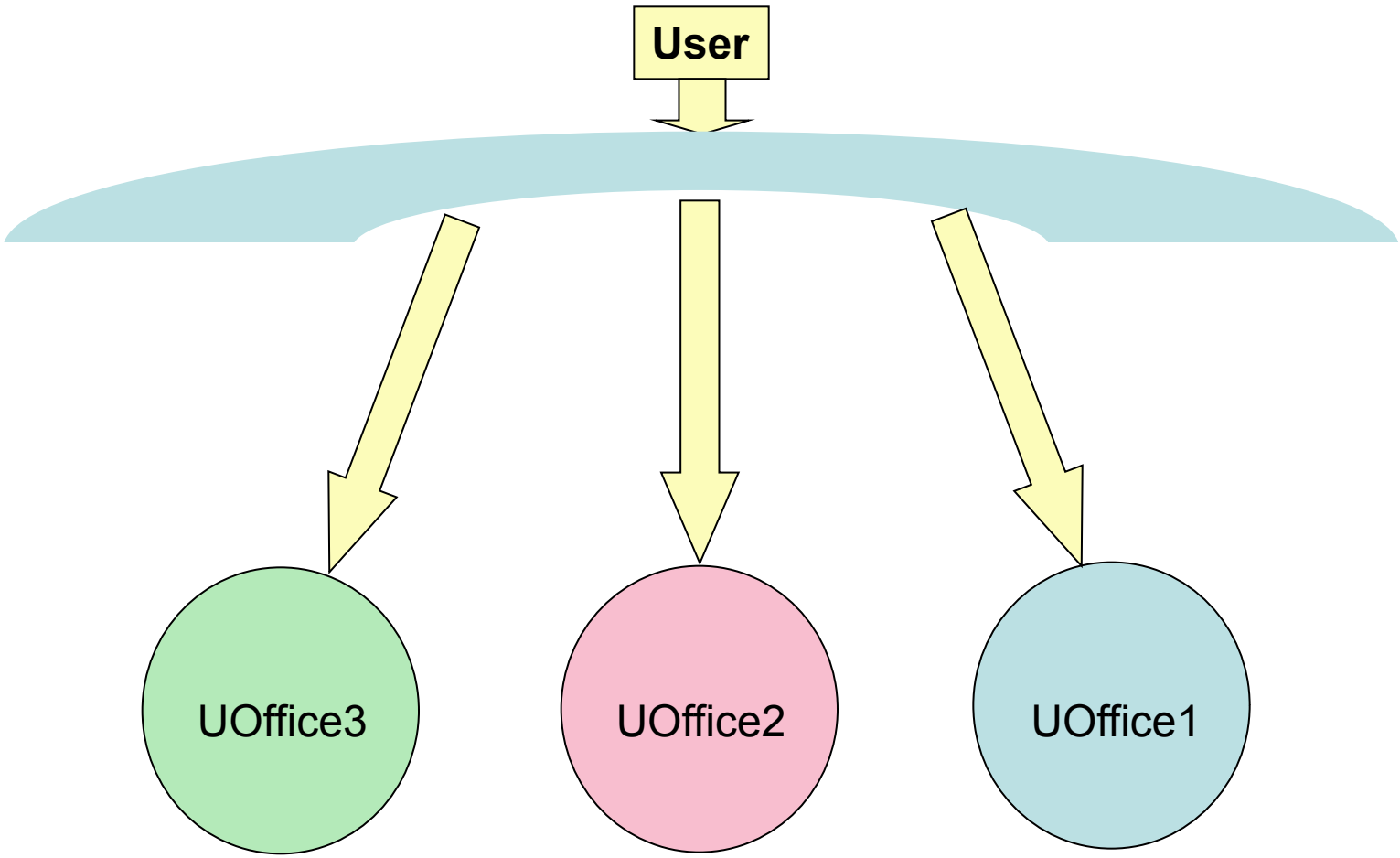
What are the IT requests?

- ***Huge datasets***
  - ➢ Novel 2D detectors, real quantum leap in data quality, but also data volumes
  - ✓ multi-image techniques (tomography, lens-less imaging)
  - ✓ molecular movies at FELs
  - ✓ 'Petabyte' becomes a 'normal' unity; time over for hard-disk in the trouser pocket
- ***Trans-facility experiments***
  - ✓ Standardize proposal procedures on EU scale
- ***Remote data access***
  - ✓ analyze data remotely at facility
  - ✓ combine datasets taken at different facilities
  - ✓ clouds (commercial, community-centered)
- ***Remote experiment access***
  - ✓ basic: passive online access to measured data
  - ✓ advanced: active control
- ***PR Issues***
  - ✓ Improve corporate identity
  - ✓ Improve public lobbying

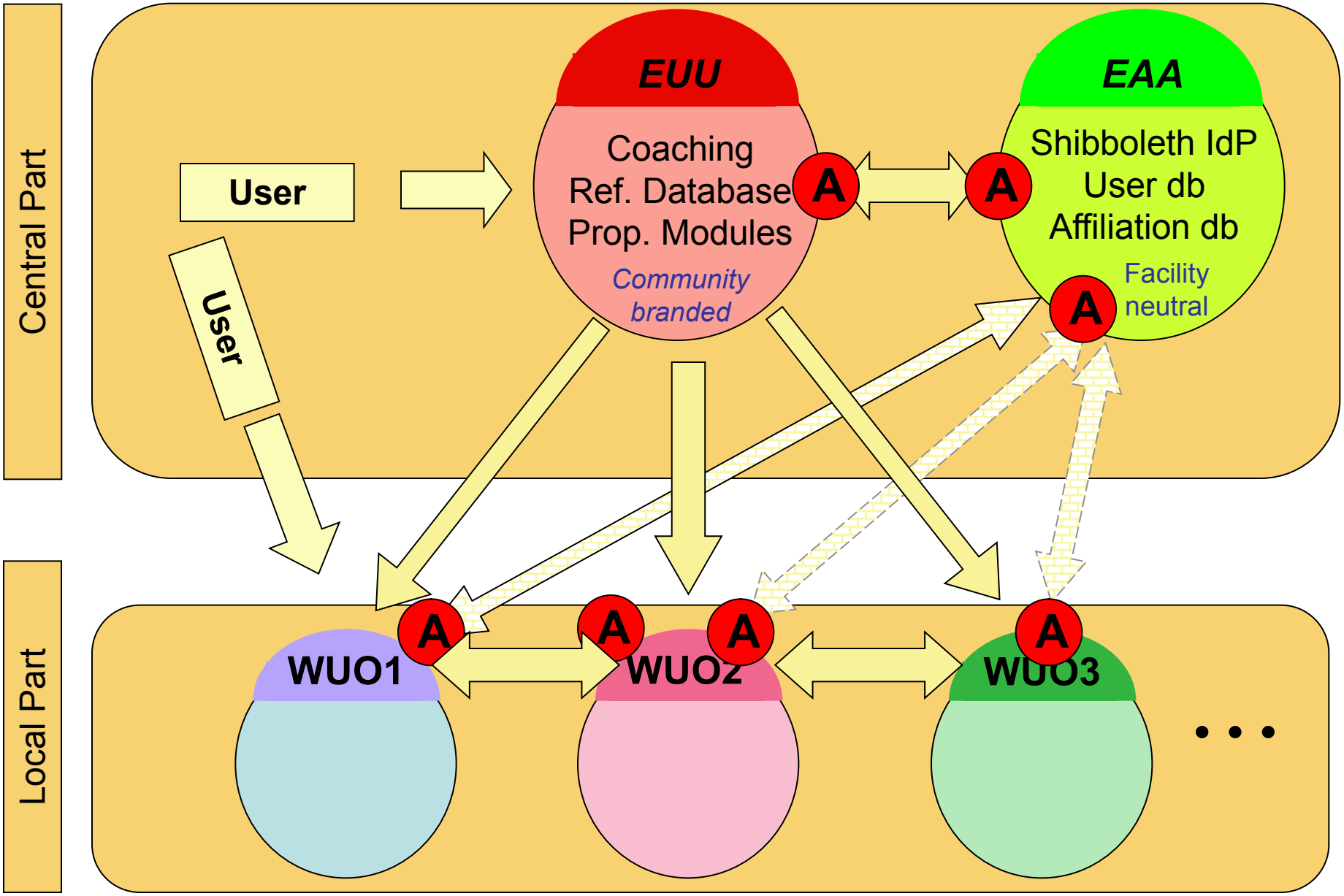
Required Solution Characteristics

- ***Incorporate confidentiality aspects***
  - ✓ High competition, especially structural biology
  - ✓ Time-window structured access to experiments and data
- ***Rely on existing local user office structure***
  - ✓ Great experience
  - ✓ Distributed operation
    - Users: manage their personal entries
    - User offices: supervising; manage authorizations
- ***Base system on professional authentication standard***
  - ✓ Shibboleth, federated Single-Sign-On System (SAML), widely used in the academic world;
  - ✓ special photon / neutron user federation
  - ✓ only one identity provider
  - ✓ supervising by local User Offices
- ***Umbrella concept***
  - ✓ Unique user identification on EU scale
  - ✓ Hybrid information storage
  - ✓ No cross-facility information exchange
  - ✓ Multi-level identification (maximum autonomy to facilities)

The Umbrella Concept



Umbrella


Central Part
User
User
EUU
Coaching
Ref. Database
Prop. Modules
Community branded
A
A
EAA
Shibboleth IdP
User db
Affiliation db
Facility neutral
A
Local Part
A
A
A
A
WUO1
WUO2
WUO3
• • •

Hybrid approach, central vs. local

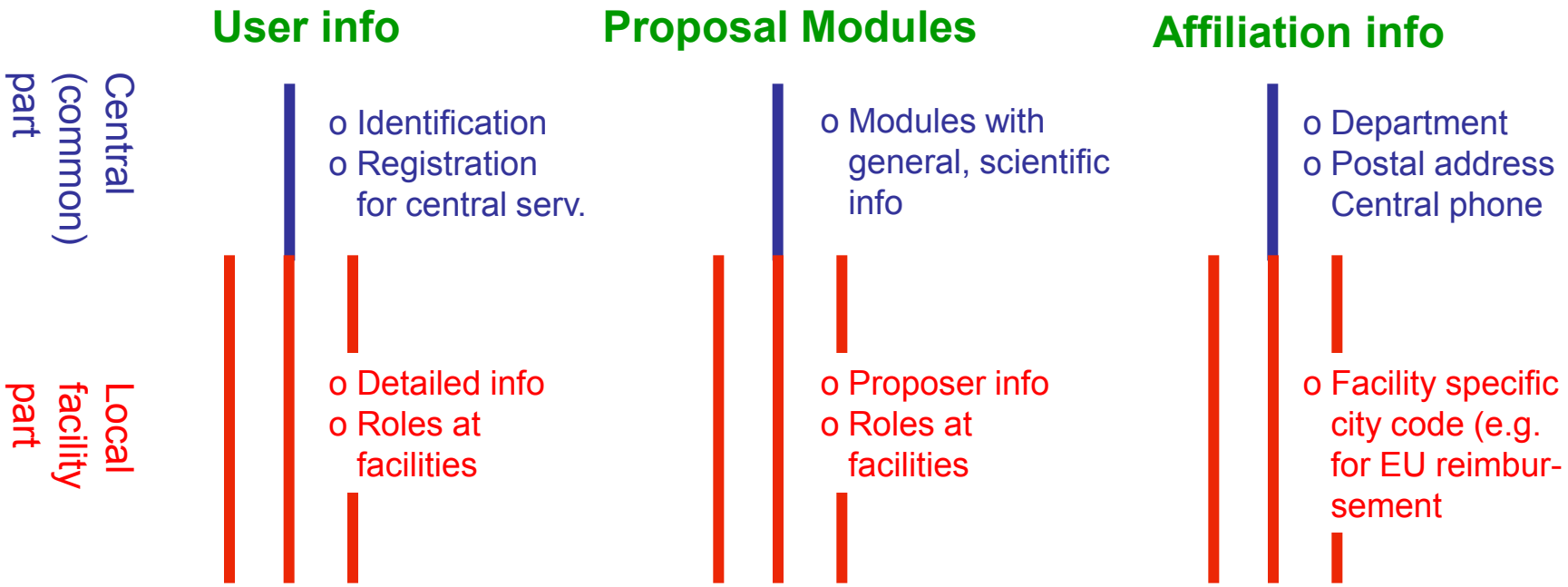
- *Central: Authentication, Unique EU-wide identification*
- *Central: Only ID-relevant info stored centrally*
- *Central: Common access portal*
- *Central: Update of user info at one place*

- *Facility-local: proposal storage*
- *Facility-local: local authorization issues*
- *Facility-local: storage of experimental data*

Hybrid character (central vs. Federated)

Answer to conflicting requests:

- Efficient technology
- Confidentiality
- Consequent distinction of authentication and authorisation



Umbrella elements

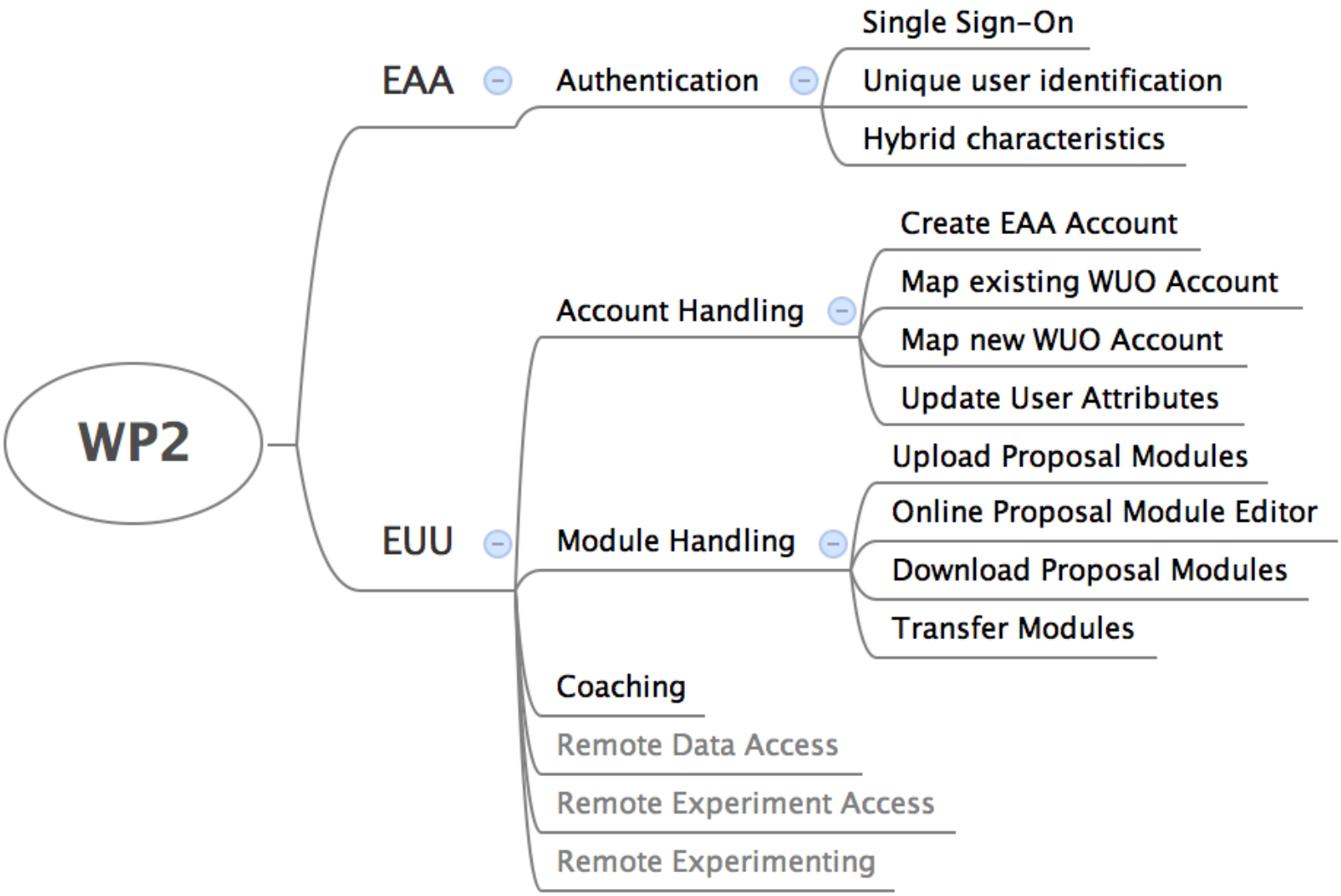
EuroFEL Umbrella prototype

- Authentication (EU-unique identification)
  - *Proposal handling* (thousands of proposals / year)
  - Coaching (support of novice users)

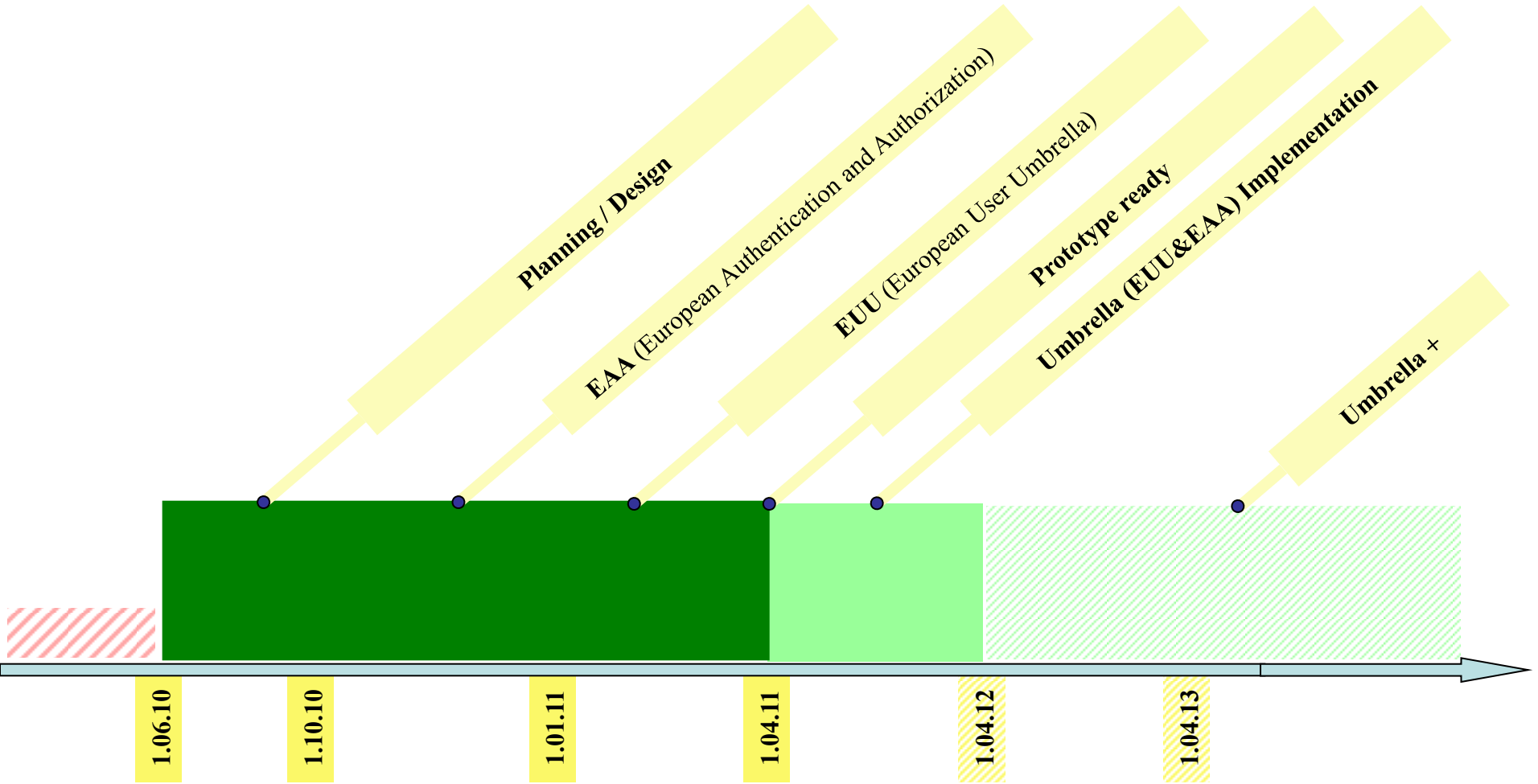
Next generation

- *Remote experiment login* (young scientists; Fedex-style experiments)
  - But more than just authentication (e.g. fire wall, experiment standardization, component *protocols …)*
- *Remote data access* (petabytes of data)
  - But more than just authentication (e.g. data format, catalogues …)

Umbrella architecture



Umbrella roadmap



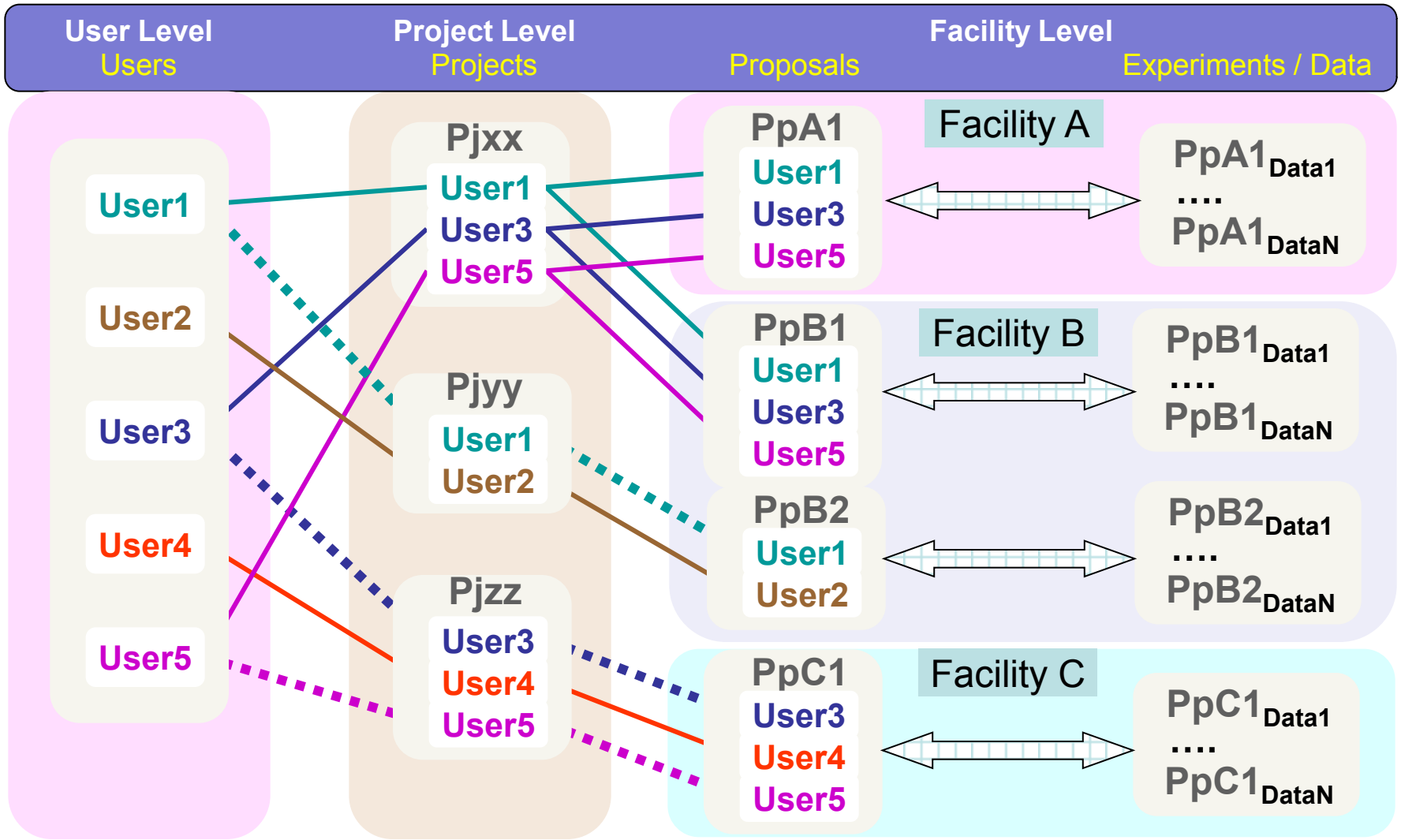
Remote data access

- **Central data storage**
  - Commercial cloud?, Bandwidth, security, costs
  - Community cloud? Bandwidth, costs? Who operates it?
  - **Keep data at sources**
- **Increased need for common science-political visibility (funds)**
  - Lobbying
  - Common web-portal
- **Cooperation between facilities**
  - Competition vs. cooperation
  - Very similar problems, exploit synergies

Remote data access, concept proposed

- **Embargo vs. post-embargo period**
  - Here only embargo (most critical, confidentiality)
- **Standard access rights rule**
  - No chance for manual central authorization
  - 1‘000s of experiments, 10‘000s of users
- **Identity by Umbrella**
  - Unique, EU-wide user authentication
- **Keep Role of proposal as organising element**
  - Users convene for a short time slot for performing an experiment
  - Principal investigator / main proposer
  - Who participates in experiment, has access right to data
  - Proposal officially accepted by facility, PI is official contact
  - PI defines who participates in the experiment (practically existing WUO tool)

Umbrella access right control



Umbrella collaborators

- **DESY, Hamburg**
  - Frank Schluenzen, Rolf Treusch
- **Fermi/Elettra, Trieste**
  - Ornela Degiacomo, Giorgio Paolucci
- **ESRF, Grenoble**
  - Rudolf Dimper, Dominique Porte, Stefan Schulze
- **HZB, Berlin**
  - Dietmar Herrendoerfer, Olaf Schwarzkopf
- **IPJ, Otwock-Swierk, Poland**
  - Robert Nietubic
- **MaxLAB, Lund**
  - Ulf Johansson
- **PSI, Villigen PSI**
  - Bjoern Abt, Stephan Egli, Stefan Janssen, Markus Knecht, Mirjam van Daalen
- **Soleil, Gif sur Yvette**
  - Frederique Fraissard
- **STFC, Didcot, Oxfordshire**
  - Anthony Gleeson

FP7 Programs, Job Sharing

- **EuroFEL WP2**
  - Prototype developments for FEL facilities (March 2011)
    - Authentication: unique user ID
    - Umbrella proposal system
- **CRISP WP6A**
  - **PSI** + ESRF, ESS, GSI, ILL, EU-XFEL
  - Authentication for management of local and remote access to facilities, experiments, data, and IT resources
  - Prototype development
- **CRISP WP6B**
  - **ESRF** + ILL, CERN, DESY +
  - Metadata management and mining service; data continuum
  - Dual local / Umbrella operation possible
- **CRISP WP6C**
  - **EU-XFEL** + DESY, ESRF, ILL +
  - High-speed Recording of Data
- **PaN-Data**
  - **PSI** + almost all European Photon / Neutron facilities
  - Authentication implementation for Photon / Neutron facilities

Conclusion

- **Increased access to facilities by non-classic users**
  - User friendliness
  - Coaching
  - Facility friendliness
- **Huge data rates for acquisition, transfer, storage**
  - Central identification
  - Remote data and experiment access tools
  - Umbrella: Tools independent from local tools
- **Increased need for common science-political visibility (funds)**
  - Lobbying
  - Common web-portal
- **Strong need for cooperation**
  - Limited awareness at top management level
  - Competition and cooperation
  - Very similar demands at all facilities, exploit synergies

Thank you for your attention!

Status and Outlook

(June 2011)

- Architecture document + road map for prototype ready
- Start development of 1st- generation Umbrella prototype
  - Shibboleth
  - deadline March 31, 2011
- Discussion 2nd-generation Umbrella (remote functionalities)
  - **'Actors':**
    - *PaN-Data*
    - *EuroFEL*
    - *ESFRI-Cluster*
    - *HDRI Helmholtz*
  - ***Tools:***
    - *GRID?*
    - *Specific development?*
  - ***Type:***
    - *Facility-friendly + user-friendly*
    - *Two-level?*
      - **Slim, simple**
      - **Strong, full-beauty IT**