

SUMMARY

We propose to hold a CECAM/Psi-k workshop devoted to the use of data representation tools (basically XML and related technologies) in atomistic simulations, with an emphasis on the needs of the electronic structure community in regard to interoperability of codes and ease of post-processing of results.

The idea has been circulating lately among an informal group of people who are already exploring, and in some cases implementing successfully, those concepts. We feel that a lot can be gained by the community through a meeting in which:

1. Experiences can be shared.
2. The available tools and approaches can be critically presented.
3. A roadmap for key developments to ease the processing and interchange of data can be discussed.

FORMAT

We propose a three-day workshop, tentatively for late January 2006 (we are working with Jan 23-25 in the CECAM application).

Number of talks: Around 15 talks and mini-talks, plus a tutorial session, and two moderated discussion sessions.

Number of participants: 15-20 people directly involved in talks and moderation of discussions, another 10 or so key people to provide feedback and opinions, and 10-15 open slots (up to a total of 40).

TITLE

Data representation and code interoperability for computational materials physics and chemistry.

OUTLINE OF THE STRUCTURE OF THE WORKSHOP

(A list of the prospective active participants whose names have been pencilled in follows at the end)

INTRODUCTION

- Goals of the workshop --- Brief introduction on the philosophy (Ph. Couch, A. Garcia)
- Introduction to data representation (data models, metadata) --- (CCLRC group)
- Brief introduction to XML and related tools (Xavier Gonze)

SHOWCASE OF CURRENT EXPERIENCES

(Without too many technical details at this point)

- Peter Murray-Rust on chemistry workflows (and possibly automatic publishing)
- Kim Baldridge and/or T. Windus "" ""
- Toby White/Jon Wakelin/Richard Bruin on SIESTA/DLPOLY/GULP output markup
- Thomas Schulthess on the ORNL efforts for interoperability of a variety of codes.
- Matthias Troyer on the ALPS framework experience
- Francois Gygi on an example of XML use in plane-wave codes.
- ...

DATA MODELS

- Peter Murray-Rust: CML as an example of domain categorization (including mentions of ontologies and dictionaries)
- CCLRC/Bristol: data models for chemistry.
- Data models for electronic-structure codes:
 - ORNL schemas,
 - Common pseudopotential format (J. Junquera, J.J. Mortensen)
- Metadata (CCLRC group)

TOOLS

- Parsers and interfaces to codes (Michael Summers (ORNL), LLNL people, A. Garcia)
- XML and legacy codes (A. Kozhenikov)
- XSLT and post-processing (including discussion of benefits of just adding markup to code output): (ALPS people, Cambridge people).
- Tools for metadata and ontology handling (tentative -- CCLRC group)

TUTORIAL SESSION

A hands-on session for users to try some of the tools, ideally taking the first steps to plug them into their own codes. (Several of the active participants will contribute to this session with examples and demos of their software, and with documentation. It would be good to have a pre-workshop distribution of some documentation)

DISCUSSIONS

We plan to hold two open discussion sessions. (Two and not just one to try to avoid the danger of a single session being sidetracked by some point. Allowing some time in between (more talks, further exchanges) will make it easier to assimilate the issues.)

A) Philosophical issues and general views of the community:

Roundtable/discussion with people such as T. Schulthess, Ph. Couch, Bjorn Winkler, Martin Dove, Xavier Gonze, F. Gygi, etc. (i.e., practitioners and/or community leaders).

Moderator: Xavier Gonze.

Agenda:

- Identification of most promising techniques.
- How to respect the needs of each (sub)-community while keeping an overall momentum.
- How to avoid blind-alleys and over-management of efforts.
- ...

(We need to make sure that key people sign up for the workshop, even if they do not have anything to present, so that their opinions are heard. "Key" means community leaders and people involved with major code development.)

A few "canned" mini-talks summarizing the steps needed to achieve a given goal will serve to focus the discussions. For example:

- Just automatization of post-processing:
 - Mark up your output with a xml-writer library.
 - Write a few xslt scripts(Jon Wakelin, Toby White)
- Share tools for post-processing with other groups:
 - Agree on /adapt to an output ontology
 - Mark your output
 - Share (or add to) the existing tools for post-processing.
- Interoperate with other codes:
 - Agree on /adapt to an input and output basic ontology
 - Use external parsers or an API to process the common input format.
 - Mark your output to conform to the common output format.
 - ...(ONRL people)
-

B) Desirability of common goals, and subsequent plans for followups and dissemination of information.

Moderators: Xavier Gonze and Phil Couch.

A "second edition" of the open discussion which might (or might not) result in the:

- Identification of common goals
- (Non-binding) recommendation of certain practices/tools
- ...

and (if goals crystallized from the discussion) a concrete plan (further workshops, web-based exchangers, etc) to keep the initiative alive.

TENTATIVE LIST OF PARTICIPANTS:

Organizers:

Alberto Garcia (Univ. del Pais Vasco, Bilbao, Spain)
Phil Couch (CCLRC, Daresbury Lab, UK)
Thomas Schulthess (ORNL, USA)

Tentative list of presenters (apart from the organizers):

Everybody on the list has been contacted in the past days, but not everybody has replied yet. Of those replying, all have expressed a strong interest in participating in the workshop.

Peter Murray-Rust (Cambridge Univ, UK)	Confirmed
Xavier Gonze (Univ. Cat. de Louvain, Belgium)	
Paul Sherwood (CCLRC, Daresbury Lab, UK)	Confirmed
Peter Knowles (Univ. of Bristol, UK)	Confirmed
Toby White (Cambridge Univ, UK)	Confirmed
Richard Bruin (Cambridge Univ, UK)	Confirmed
Jon Wakelin (Univ. of Bristol, UK)	Confirmed
Kim Baldrige (Univ. of Zurich, Switzerland)	Confirmed
Theresa Windus (PNNL, USA)	
Michael Summers (ORNL, USA)	Confirmed
Anton Kozhenikov (Ural Branch, Acad. of Sciences, Russia)	
Francois Gygi (LLNL, USA)	Confirmed
Matthias Troyer (ETH Zurich, Switzerland)	Confirmed
Jens J. Mortensen (Tech. Univ. of Denmark)	
Javier Junquera (Univ. de Cantabria, Spain)	Confirmed

Other people that should come to participate in the discussions

(due to their positions, or their participation in ongoing efforts, or as members of code development teams)

Paolo Giannozzi (Univ. of Pisa, Italy -- Quantum Espresso development)	Confirmed
Martin Fuchs (Fritz Haber Inst, Germany -- FHI code development)	
Peter Blaha (Tech. Univ. Vienna, Austria -- WIEN code development)	Confirmed
Georg Kresse (Univ. of Vienna, Austria -- VASP code development)	
Emilio Artacho (Cambridge Univ., UK -- Siesta development)	Confirmed
Martin Dove (Cambridge Univ., UK -- eMinerals)	Confirmed
Peter Nielaba (Univ. Konstanz, Germany -- chairman of MolSimu network)	
Walter Temmerman (Daresbury, Psi-k management)	
Bjorn Winkler (Univ. Frankfurt, Germany)	

...

Other participants:

Phil Couch organized a workshop on data representation in quantum chemistry in April 2004 which was attended by people from a variety of fields. Some of them might want to participate in the proposed workshop.

We should keep a working limit of 40 attendees.

BUDGET

For a three day workshop, the CECAM automated budget estimator gives a magic figure of 8000 EUR, which might be enough to cover the room and board expenses of the participants. We would like, if possible, to cover the travel expenses of the speakers/presenters. We might be able to use american funds for the american participants (from DOE, as they are mostly in government laboratories) but we might need another 5000-6000 EUR or so from Psi-k to cover the europeans.